Vol. 53, No. 3 May, 1946

Psychological Review

EDITED BY
HERBERT S. LANGFELD
PRINCETON UNIVERSITY

CONTENTS

Walter Bradford Cannon: 1871–1945: ROBERT M. YERKES 137

The Ability to Learn: HERBERT WOODROW 147

A Critique of 'Neutral' Items in Attitude Scales Constructed by the Method of Equal Appearing Intervals: ALLEN L. EDWARDS 159

Some Consequences for History and Psychology of Langmuir's Concept of Convergence and Divergence of Phenomena: IVAN D. LONDON .. 170

A Note on the Similarities and Differences Between Projective Tests and Ability Tests: DANIEL J. LEVINSON 189

PUBLISHED BI-MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1227 NINETEENTH STREET, N. W., WASHINGTON 6, D. C.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879.

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears bi-monthly beginning in January. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave*, FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears bi-monthly (beginning 1944), one volume per year, each volume of six numbers containing about 520 pages.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 350 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 400 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

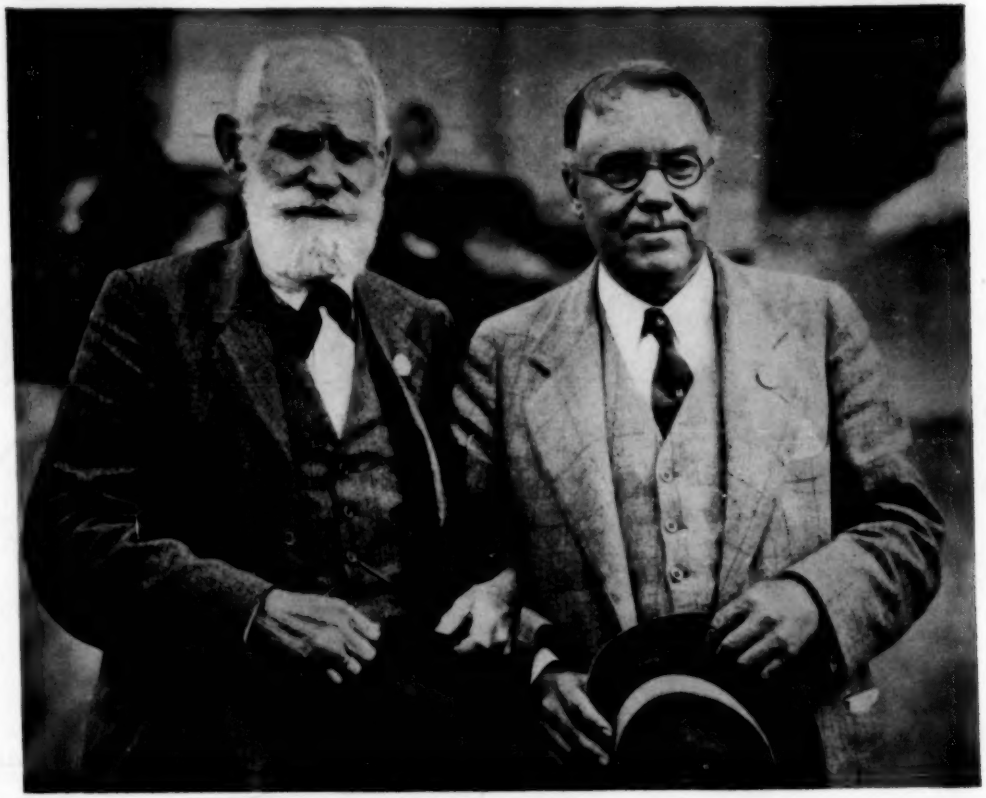
DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1227 NINETEENTH STREET, N. W., WASHINGTON 6, D. C.



J. Pawlow Walter B. Cannon

VOL. 53, NO. 3 MAY, 1946

THE PSYCHOLOGICAL REVIEW

WALTER BRADFORD CANNON

1871–1945

"Rarely does a man of science have opportunity to devote himself to productive scholarship for half a century, and still more rarely does he continue his devotion if he has that opportunity. And yet, for nearly sixty years Pavlov was an outstanding productive scholar. During that long period he dedicated his talents and his tireless energy to the revelation of new and significant facts in the functioning of higher organisms. With incomparable zeal and enthusiasm, and with admirable singleness of aim, he spent his life seeking the explanation of the natural processes of living creatures. As a result of his ingenious methods and extraordinary technical skill, our knowledge of the functions of the digestive glands and the functions of the more complex processes of the nervous system has been largely transformed. For indefinite time Pavlov will remain a shining example of Russian genius" (6, p. 5).

Thus appreciatively Walter Cannon wrote of his friend and colleague Ivan Petrovitch Pavlov. Only a few words need be changed to render this tribute perfectly applicable to its writer. For both men were preëminent in their profession and extraordinarily useful as scientists and as citizens. With intent to suggest their intimacy and the striking similarities of their interests and scientific careers, the portrait selected to accompany this biographical sketch presents both Cannon and Pavlov. It was taken in Boston in 1930.

Walter Cannon was born at Prairie du Chien, Wisconsin, October 19, 1871. He died at his summer home, Franklin, New Hampshire, October 1, 1945. He was well-born of able, energetic, socially-minded parents. His father was serious-minded, studious, and ingenious. Although engaged in the railroad business all his life, he was especially interested in farming and medicine and read extensively on those subjects. It was his earnest hope that his son might become a physician.

Following a profitable public school education in Milwaukee and St. Paul, Walter was encouraged by one of his high school teachers to seek collegiate and professional training in the east. It was a career-determining adventure. He entered Harvard College with approximately $180, which his father had been able to give him as capital, and thereupon he 'worked his way' during eight years of collegiate and medical education. Upon completion of his medical course, he was offered an instructorship in physiology at the Medical School and a like appointment in the College department of zoölogy. He accepted the former opportunity and began in 1900 a career as teacher and investigator in Harvard which continued until 1942. Acting promptly on the conviction that happiness is born a twin, he married Cornelia James. Thus began an ideal partnership in which the natural caution and conservatism of a highly conscientious and rigorously trained man was invaluably complemented and supplemented by the fearless liberalism, independence, and in-

itiative of an extraordinarily energetic, able, and well-educated woman. For nearly fifty years Cornelia James was the greatest personal influence in Walter Cannon's life and the most beneficent. Theirs was a rarely happy and successful marriage of mutual affection, admiration, and helpfulness. Children and grandchildren rise to call it and them blessed.[1]

Of excellent constitution and notably vigorous both physically and mentally, Cannon led an active, healthfully varied, and always planful life. In the light of its fruits, one may adjudge him fortunate in the prime essentials for successful living. Certainly this is true of heredity, early home environment, educational opportunities, vocational choice, occupational opportunities and conditions, and social relations.

Early he acquired habits of work and discovered the values of care and thoroughness. Conspicuous among his traits were industry, resourcefulness, honesty, friendliness, unselfishness, and never-failing motivation for achievement. His zeal and enthusiasm for research and his tireless devotion to his labors as experimental physiologist led him to discoveries which mankind holds precious.

But in the molding of Walter Cannon's personality, heredity, social tradition, educational advantages, and professional experiences were importantly supplemented by a multitude of coöperative human contacts and intimate friendships. For he was naturally a kind and generous man who gave himself as unstintingly to his friends as to his work.

Cannon's philosophy of life, nowhere explicitly and completely presented, may be inferred from his activities and from fragmentary statements. What he lived by and for is readily discoverable: his compelling faith is manifest, even as a youth. Born into a strict Calvinistic family, he early broke with religious authoritarianism and achieved independence of dogma and 'revelation.' In later life he stated his conviction that natural events, and in particular those which constitute life, make for useful ends; that bodily processes appear to be purposive. Likewise he asserted that structure and function, body and mind, are inseparables in an organismic unity. Naturalism would seem an appropriate label for his philosophy of life, for he asserts that in the study of organic functions, explanation does not require us to assume the supernatural or the extranatural. In accordance with these articles of faith, his studies of man involved the assumption that the knowledge and understanding to be achieved by research should have ethical, moral, and other social values and implications. It is evident, then, that in all of his research he was striving for knowledge which would enable man to live more intelligently, more wisely, and more happily, and was motivated by the conviction that such knowledge is essential to wisdom and righteousness.

His career in research was consistent, logical, and except for the interruptions of two wars, continuous. As an undergraduate, under the guidance of Charles B. Davenport, he inquired into the influence of light on the direction and rate and movement of organisms. This, his introduction to original work, doubtless helped to set the pattern for his professional life as creative scholar, teacher, and investigator. In the years which followed, several areas or foci of research-interest and contribution are evident. They are not sharply demarcated, for the earlier interests tended naturally to continue as others were added.

[1] It is impossible to document this biography adequately, for it is based almost as largely on personal and professional contacts, over more than forty years, as on available scientific and autobiographic reports. Under the circumstances it is difficult to write objectively and impossible to write impersonally.

First in time, and ranked high in importance by physiologists, physicians, and surgeons, is his study of the movements of the stomach and intestines and of their significance and relation to other gastro-intestinal functions. For this series of experimental investigations, Cannon ingeniously availed himself of the recently discovered X-ray as instrumentality for direct observation. In accordance with an ideal of public service which was exemplified throughout his life, he presently made available to the general reader authoritative and practically useful accounts of his studies of digestive processes, their relations and conditions. There are two books, of which the first (1) gives an account of the mechanical factors of digestion, and the second (7), of the nature of digestion and its relation to health.

The farther he proceeded in his studies of the physiology of digestion, the more inclusive became his interest. Gradually, from mechanical factors, attention shifted to the rôle of endocrine secretions, and in the course of prolonged work signal contributions were made to knowledge of the functions of the adrenal and thyroid glands and the effects of their secretions. Furthermore, it is evident from his writings that while observing gastro-intestinal activities and the influence of endocrine secretions, he repeatedly noted and was deeply impressed by evidences that many bodily processes vary in relation to the affective state of the subject. For him, presently, 'bodily effects of emotions' came to designate a problem-complex of compelling interest and importance.

The three foci of interest which have been listed filled Cannon's years of research prior to World War I. And the major results of nearly two decades of strenuous effort, initially presented in technical reports, are summarized and interpreted for the non-specialist, and as well for the lay reader, in a widely read and extremely valuable book on the relations of bodily events to emotions (4).

In 1917 military need and opportunity diverted him from his current research activities, and throughout the war his experimental skills and practical resourcefulness and wisdom were devoted largely to the study of surgical shock under war conditions, and to the search for effective measures of prevention and control (2).

Upon return from military duty he concentrated his attention increasingly, for several years, on a systematic study of the regulation and stabilization of such essential bodily conditions and processes as the constitution of the blood and other body fluids. This led him to inquiries into neural functions, which finally convinced him that the sympathetic nervous system is primarily a regulative mechanism which maintains what he designates as bodily 'homeostasis' and thus protects the organism from dangerous or injurious rate or degree of change in substances or functions whose stability and constancy, within certain limits, are essential to normal living. Concerning this field of discovery, he writes as follows:

> "The steadiness of the 'milieu interne,' therefore, may be regarded as an arrangement of economy. And the course of evolution of higher organisms has been characterized by a gradually increasing control of the functions of that *milieu* as an environmental and conditioning agency. Just insofar as this control has been perfected, both internal and external limitations on freedom of action have been removed and risks of serious damage or death have been minimized. The central problem in understanding the nature of the remarkable stability of our bodies, therefore, is that of knowing how the uniformity of the fluid matrix is preserved" (5, pp. 269–270).

Presumably Cannon ranked this body of observations and the conclusions to which they led him as most important among his contributions to physiology and medicine. It is, indeed, a stirring panorama of fact and principle which unfolds to the reader as he follows these studies of organic regulation. For the convenience of those who cannot or will not consult original sources, the outcome of these years of research and reflection is available in a book on homeostasis which the author entitled *The Wisdom of the Body* (5).

Again a shift of research interest occurs and we find the master physiologist giving himself whole-heartedly to searching out the secrets of neural processes, with intent to discover the nature of the nerve impulse and the conditions essential to its proper functioning. This he has himself described as inquiry into the 'chemical mediation of nerve impulses.' It was in the course of this series of investigations that he discovered sympathin, a substance appearing in the blood stream as a result of the action of sympathetic nerve impulses on certain bodily structures.

Unfortunately his great adventure in neuro-physiology was partially interrupted by World War II. As Chairman of the National Research Council's Committee on Shock and Transfusions, he was enabled to make available to military authorities and other professional colleagues the knowledge and wisdom which he had acquired in his previous experience with war-time surgical shock, and to take a leading part in the development and perfecting of new measures for shock therapy. The story of the use of blood and blood substitutes for emergency treatment during the war is familiar to all.

In his remaining years, although struggling against pathological changes which defied medical skills, he continued his neurological studies with reference especially to the 'effects of severance of nerves on sensitizing isolated structural elements to chemical agents.' His span of life did not permit the rounding out of his chemical mediation and related neurological studies and the preparation of an interpretative general account of what he had discovered and assimilated to his science of life. The nearest approach to such a book is his autobiographic account of a scientist's experiences in medical research (10).

Numerous adventures of Cannon the physiologist into the research no-man's-land between physiology and psychology challenged the attention of psychologists and stimulated much appreciative and adverse critical discussion of his discoveries and conclusions. The evidences indicate that he was not wholly at home in psychology; that his thorough knowledge of anatomy and physiology was not matched by adequate familiarity with the spirit and objectives, the emerging methodologies, discoveries and generalizations of twentieth century experimental psychology. Whereas Pavlov, while striving to be a consistent objectivist, almost defiantly denied scientific status to psychology and ignored psychologists, Cannon accepted the subjective and objective as aspects of observable phenomena, as witness the following, "If we agree that we are organized as a psycho-organismic unity we need not hesitate to use any convenient terms in mentioning any aspect of behavior or experience, for then we understand that these terms designate one aspect or the other of a living, systematic whole" (8, p. 1464). It appears from the sum of their labors that Pavlov's was the simpler, safer way, whatever its logical and pragmatic defects, for the course which Cannon chose led him into misleading inconsistencies of statement and seeming self-contradiction.

The following exhibits are offered in

partial justification of the previous general statements. Cannon asserts that 'consciousness' and 'mind' have the same meaning; that the 'emotion' may function as 'stimulus' (from which it may be inferred that the psychic factor is effective). These statements from an address to his medical colleagues are *apropos*. "Fears and worries, persistent hatred and resentment—what pathology have they? Are they not manifestations of the psyche? And who knows about the psyche?" (8, p. 1455).

Adducing the Weber-Fechner law as an example of the "error of unwarranted conclusions," Cannon quotes, in support of his action, the opinion of William James that the probable psychological consequence of Fechner's law "is just *nothing*" (10, p. 127). At this point the psychologist-reader may wonder whether Cannon knew that near the end of his professional career the gifted introspective expositor and interpreter of psychology, William James, expressed also the opinion that the principal fruits of experimental psychology to date were of little moment.

It was inevitable that Cannon's sustained and extremely fruitful inquiries into the functions of the nervous system in their relations to the 'bodily effects of the emotions,' eventually should stir his critical interest in the James-Lange theory of emotion. These men had stated hypothetically that the 'simply apprehended' situation becomes 'emotionally felt' only by reason of the organic sensations which are induced by the situation and the organism's response to it. To Cannon it appeared that the experimental and clinical evidences were so unfavorable to the organic sensation theory that it must be rejected. He points to the following among other categories of fact which James and Lange appear to have overlooked.

Even total separation of the visceral organs from the central nervous system does not essentially change emotional expression. Furthermore, in connection with quite different emotions and in non-emotional states, like visceral changes are observable. The internal organs are relatively insensitive structures and in general visceral changes occur too slowly to be an important source of emotional experience. Finally, the artificial induction by adrenin of visceral changes observable in emotions does not produce the emotions.

From the observations reported by other investigators, notably Head (12) and Dana (11), and his own experimental findings, Cannon noted the rôle of the hypothalamus in the control of certain emotional expressions. He thus expresses his early conviction, "All these observations, experimental and clinical, consistently point to the optic thalamus as a region in which resides the neural organization for the different emotional expressions" (4, p. 364), What may appropriately be designated the Head-Cannon thalamic theory of emotion has been more explicitly stated and more adequately defended by Cannon than by anyone else. The essence of the theory appears in the following sentences from Cannon:

"In other words, for the theory that emotional experiences arise from changes in effector organs is substituted the idea that they are produced by unusual and powerful influences emerging from the region of the thalamus and affecting various systems of cortical neurones. This view accords with the pertinent physiological facts now available. It can be applied to the 'subtler' emotions, which the James-Lange theory had difficulty in explaining. It offers interesting suggestions for the study of emotional expressions experimentally in lower animals and in human beings under the influence of various drugs and in pathological states" (3, p. 267).

This theory involves the following important assumptions for which, in Can-

non's opinion, there is observational basis. Afferent impulses due to stimuli which ordinarily arouse emotion pass to the cortex where they are integrated to induce appropriate behavior. Simultaneously a portion of the hypothalamus is excited either as the sensory impulses are relayed or secondarily by impulses from the cerebral cortex. Thereupon the hypothetical thalamic center for control of emotion discharges precipitantly, thus exciting effects which are responsible for the emotional expression and also initiating impulses to the cortex, where the emotional quality is added to the sensations resulting from the emotional situation. Furthermore, the impulses from the thalamus are supposed to account for emotional tension and the motivational character of the experience.

Contrary to the assumption of some of his critics, Cannon obviously does not consider the thalamus the seat of emotional experience. Instead it is a contributory neural mechanism upon which the activation of inborn patterns of emotional expression depends.

Some ten years after Cannon's original presentation and defense of the thalamic hypothesis, Lashley systematically reviewed the neurophysiological evidences which bear on it. His succinctly stated conclusions follow:

"The supposed evidence that the thalamus adds the affective or emotional character to sensations breaks down completely when subjected to critical analysis. The affective changes resulting from thalamic lesion are restricted to a small group of somesthetic sensations and cannot be interpreted as a general change in affectivity. The changes correlate definitely with the special properties of conduction, summation, and irradiation of this group of sensory processes and not at all with a specific locus in the thalamus. The pathological changes following thalamic lesions are primarily in the character of the sensations, in intensity, duration, localization, and are therefore not relevant to the problem of affect. There is no evidence whatever that the thalamus contributes facilitative impulses which might form a basis for the motivational aspects of emotion. Thus, the only part of the thalamic theory of emotion which has factual support is the localization of motor centers for emotional expression within the hypothalamus. It seems certain that these motor centers do not contribute directly to other aspects of emotion and there is no evidence for the existence of other affective or emotional centers" (13, p. 60).

From our present vantage point, the inadequacy of the James-Lange theory of emotion is evident, while the adequacy of the Head-Cannon thalamic theory is questioned by competent specialists. Doubtless had James, Lange, Head, and Cannon been able to sit down together to review critically and to correlate and evaluate the pertinent evidences now available from morphological, physiological, and psychological sciences, they would have produced a much more nearly adequate theory in replacement of those previously proposed. Theorizing relative to the nature and mechanisms of affective behavior and experience is at best discouraging, for the march of discovery is rapid, though irregular, and our understanding of the mechanisms of behavior-experience today may be outmoded tomorrow!

But whatever the fate of the theory of emotions to the defense of which Cannon devoted so much effort, the significance of his factual contributions to our general understanding of the rôle of the sympathetic nervous system and of the relations of emotional expression and experience to other physical and mental phenomena, is clear. By and large, fact is more important than theory. As a master craftsman and a leader in the 'proper study of mankind,' Cannon takes his place among the great.

Sketches of man, scientist, citizen, should blend into a portrait of the personality. In the remaining paragraphs of this biography attention will be centered on the character and activities of citizen Cannon. What has gone before may have given the impression that his most significant public service was rendered through biological research. Probably that is not true, for his services as teacher, counselor, defender of freedoms, promoter of social progress, and friend of all mankind, together seem far to outweigh the immediate values of his original discoveries.

First, the teacher. Cannon worked for his school and his students in the confidence that teaching is the most important of all callings. Forty classes of medical students passed through his classroom, laboratory, and study. In them his ideals of professional achievement, obligation and worth, of respect for the truth, and of creative scholarship are perpetuated, for today all over the world there are physicians, surgeons, investigators, administrators, who remember him with admiration and gratitude as teacher and friend. Truly, few influences in life are comparable with that of the great teacher.

Almost at the outset of his professional career in Harvard, Cannon proposed and effectively advocated the use of the case method in teaching medicine. The quality of progressiveness thus expressed was responsible for much that he did as faculty member to improve medical education and practice and to aid and abet President Eliot in converting the Harvard Medical School into a model institution for professional training.

Throughout his life Walter Cannon fought valiantly and determinedly for human freedoms. This is exemplified on the personal side by his early conflict with religious authority and dogma. Then and ever after he strove to achieve and preserve personal freedom of thought, inquiry, belief, self-development, and self-expression. This early adventure marked the beginning of a battle against superstition, wishful thinking, and timid conservatism which ended only at the end of the trail.

As youthful professor of physiology, he was made chairman of the American Medical Association's Committee for the Protection of Medical Research and in this capacity, for nearly two decades, he marshalled and directed the forces of his great profession to prevent legislative restriction of medical research. By conference, public address, publication of pertinent facts and arguments, he and his colleagues labored successfully, but at high cost of energy and time, to counteract and correct the propaganda and ill-advised restrictive proposals of that small minority of our fellow citizens who make a fetish or a religion of 'antivivisection' and who sometimes act as though other organisms are more worthy of considerate treatment and preservation from accident or disease than is man. From certain points of view it is indeed a tragedy that so much of Walter Cannon's strength should of necessity have been devoted to the defense of freedom of inquiry. Yet, even as this is being written, our daily press has carried extended quotations from his authoritative addresses on the proper use of animals in experimental inquiries. His beneficent influence therefore lives, and so also will the freedom of inquiry which he defended as essential to the amelioration of many human ills and to the improvement of man.

There is yet a third realm of freedom which engaged his courageous, resourceful, and unselfish endeavor. I refer to the freedom of social action and the guaranty of justice to those who have neither wealth nor power. Professional colleague of Dr. Juan Negrín, premier

of the Republic, Dr. Cannon espoused the cause of Republican Spain as against the action of his own country and of Britain. He served for two years as National Chairman of the Medical Bureau to Aid Spanish Democracy. The assistance rendered cannot be measured by the considerable sums raised and expended for medical personnel, equipment and supplies. Of this experience in social service, during which he was arraigned as Bolshevist, Communist, enemy of the Catholic church, he writes simply "I only regret that our efforts were not more effective in helping at last to achieve the common degree of freedom among a people long held in ignorance and economic repression" (10, p. 162).

In the service of international understanding, enlightenment, and good will he served also as Visiting Professor at the Peking Union Medical School in 1935. Here he observed the weakness of China's ancient civilization, the stirrings of a new life therein, and the arrogant encroachment of the Japanese. Later it was his privilege, as he puts it, to assist in the work of the American Bureau of Medical Aid to China and also of the United China Relief. And in the twilight of his busy life, despite encroaching physical ills, he gave his earnest support to the cultivation of Russian-American mutual understanding, friendship, and coöperation, as President of the American-Soviet Medical Society, deeming this opportunity for service to humanity through leadership an honor.

Cannon's war emergency services were as obviously those of citizen of the world and friend of mankind as of scientist and medical investigator, physician, or surgeon, for he was all of these. They have been noted with brief characterization in earlier paragraphs.

Few indeed were the years in Walter Cannon's professional life when he was not giving himself wholeheartedly to membership or leadership activities in conferences, committees, commissions, and other bodies created to organize, promote, or otherwise advance biological research. First and last these nationally and internationally significant services on behalf of human welfare and enlightenment totalled years of unselfish endeavor. It is worthy of note by our narrowly commercial culture that these services were freely given, not bought.

What seems to the writer citizen Cannon's most important public service has been reserved to the last. It is his serious proposal that the principles of organic self-regulation (biological homeostasis) be applied experimentally to social organization and regulation. In his notable work on *The Wisdom of the Body* (5) he includes as epilogue a characteristically modest discussion of relations of biological and social homeostasis. The subject was later elaborated in his address as President of the American Association for the Advancement of Science (9). Statesmen, social planners, and practical politicians would do well to read and reflect on these statements of fact and suggestions for social progress.

We are constituted of highly instable substances upon which our patterns of action and experience depend. It is essential to our survival, comfort, and freedom of action that stability of substance and process be at all times maintained, within certain safe limits of variation. In the course of organic evolution, muscular, glandular, neural, and other structural mechanisms have evolved as regulators of essential organic processes. The history of the race (phylogeny) synoptically repeats itself in the history of the individual (ontogeny). At birth the infant is only partially self-regulatory. Gradually, with growth, homeostasis appears and

shortly the individual, if normal, is adequately protected by regulatory mechanisms against the hazards of its external and internal environment. Whereas automatic regulation of functions through the evolution of special mechanisms has reached a high state of development in the human individual, in comparison with that observed in such lower vertebrates as fishes or amphibians, it is evident, as Cannon asserts, that *social organization* is relatively crude and inadequate—possibly as yet in an early stage of evolution.

What, then, Cannon asks, does the study of homeostasis in the individual suggest by way of principles applicable to the further improvement of social-economic organization and regulation? The answer may be stated in his own words.

"The organism suggests that *stability is of prime importance.* It is more important than economy. The organism throws away not only water and salts, but also sugar, if they are present in excess in the fluid matrix. This rejection is uneconomical. The organism is driven into convulsions if the sugar supply runs too low, and the convulsions mark the acme of the manoeuvres which bring forth extra sugar from the hepatic reserves to restore the normal glycemic percentage. Violent shivering may be induced to develop the additional heat which prevents a fall of body temperature. . . .

"The organism suggests, also, that there are early signs of disturbance of homeostasis which, if sought, can be found. These warning signals are little known in the social organism, and yet their discovery and the demonstration of their real value would make contributions to social science of first-rate importance. In the complexity of modern social interrelations the strategic control would appear to reside in the devices for distributing goods, in commerce and the flow of money rather than in manufacture and production. Our bodily devices would indicate that the early warning signals, pointing to social and economic danger, should perhaps be sought in sensitive indicators of fluctuations of the commercial stream, though the causes of these fluctuations may be found in industry.

"The organism suggests, furthermore, that the importance of stability warrants a specially organized control, invested by society itself with power to preserve the constancy of the fluid matrix, *i.e.*, the processes of commerce. Does not this imply that when there is prospect of social perturbation there should be power to limit the production of goods to a degree which would reasonably adjust the supply to the demand? power to lay aside stores of goods which could be released if crises arise? power to require the accumulation of wage reserves which could be used at times of temporary unemployment? power to arrange emergency employment or training for new types of labor skill? and power to accelerate or retard the routine processes of both the production of goods and their distribution, in accordance with desirable adaptations to internal or external disturbing factors? It is noteworthy that in the bodily organism such powers as storing or releasing material reserves, hastening or checking continuous processes, are exercised not by the cerebral cortex, where adaptive intelligence is mediated, but by lower centers of the brain which work in an automatic manner when appropriate signals call upon them to act" (5, pp. 299–300).

What major advantages to mankind might be expected from the intelligent application of such principles of physiological homeostasis to our social-economic system? Cannon replies, "The main service of social homeostasis would be to support bodily homeostasis. It would therefore help to release the highest activities of the nervous system for adventure and achievement. With essential needs assured, the priceless unessentials could be freely sought."

"Just as social stabilization would foster the stability, both physical and mental, of the members of the social organism, so

likewise it would foster their higher freedom, giving them serenity and leisure, which are the primary conditions for wholesome recreation, for the discovery of a satisfactory and invigorating social *milieu*, and for the discipline and enjoyment of individual aptitudes" (5, pp. 305–306).

Never in human history have suggestions for the improvement of our social institutions, particularly the governmental, and the effective regulation of their functions been more timely than are those of Cannon today. Diligent study of the principles of bodily homeostasis and of their immediate applicability may help to channel and speed the process of social evolution and immeasurably increase our debt to him whom we now mourn as friend and colleague.

REFERENCES

1. CANNON, W. B. *The mechanical factors of digestion.* London: Arnold, 1911.
2. ——. *Traumatic shock.* New York: Appleton, 1923.
3. ——. Neural organization for emotional expression. *Wittenberg Symposium.* Worcester: Clark University Press, 1928, 257–269.
4. ——. *Bodily changes in pain, hunger, fear and rage: an account of recent researches into the function of emotional excitement* (2nd ed.). New York: Appleton, 1929.
5. ——. *The wisdom of the body.* New York: Norton, 1932.
6. ——. Ivan Petrovitch Pavlov. *Research Bull. on the Soviet Union,* 1936, **1**, 5–8.
7. ——. *Digestion and health.* New York: Norton, 1936.
8. ——. The role of emotion in disease. *Ann. intern. Med.,* 1936, **9**, 1453–1465.
9. ——. The body physiologic and the body politic. *Science,* 1941, **93**, 1–10.
10. ——. *The way of an investigator: a scientist's experiences in medical research.* New York: Norton, 1945.
11. DANA, C. L. The anatomic seat of the emotions. *Arch. Neurol. & Psychiatr.,* 1921, **6**, 634–639.
12. HEAD, H. *Studies in neurology.* London: H. Frowde, Hodder & Stoughton, 1920, Vols. 1–2.
13. LASHLEY, K. S. The thalamus and emotion. PSYCHOL. REV., 1938, **45**, 42–61.

ROBERT M. YERKES
Yale University

THE ABILITY TO LEARN

BY HERBERT WOODROW

University of Illinois

I. Definitions and Generalizations

The area to be considered will be limited by using a restricting definition of learning ability. In his book on *The Psychology of Human Learning*, McGeoch states that "Learning, as we measure it, is a *change* in performance with practice" (13, p. 3). It will be noted that this statement is not specific with respect to either of two questions. It leaves open the question between what two points in the course of practice the change to be considered extends. It also fails to specify how change in performance is to be measured, that is to say, in terms of what kind of scores it is to be appraised. Though McGeoch's definition is regarded as acceptable, in order to avoid some of the complexities of this matter, one important distinction will be made, namely, one between past learning and future learning. The time taken as dividing the past from future learning is the time at which the subject's performances first become the object of experimental observation and record. Thus, in the case of a subject in a practice experiment, by past learning is meant the learning which may have occurred prior to the beginning of the experiment. It is commonly assumed, though on questionable grounds, that this past learning begins with a performance representable by a score of zero. By the future learning of a subject is meant whatever learning occurs after the time when the subject's performance is first measured by the experimenter.

This distinction between past and future learning is a necessary one unless we are to make very speculative extrapolations into the past from such data as are ordinarily obtainable. However regrettable this may be from the general scientific viewpoint, it may be pointed out that, practically, we are ordinarily interested in the question of learning ability as it will affect improvement beyond some time well along in the life of the individual, as, for example, from the time he enters a school grade, or begins the study of a new subject, or accepts a new position. The learning envisaged, then, will be limited to that which occurs beyond some time when it is feasible to obtain an initial measure of performance from which to calculate gains occurring thenceforth. It apparently is not necessary to indicate the time at which the change or improvement to be taken into consideration terminates. The relationships to be stated below probably hold for practice over any length of time, even sufficient to raise the individual to his hypothetical 'physiological limit.'

In order to make any accurate generalizations concerning learning ability it is necessary also to introduce some limitations into the types of scores by which change may be measured. The chief question here is whether the change is to be taken as simply the difference obtained by subtracting the first observed score from the last one, or as this difference considered as a proportion, or percentage, of the first score. The present discussion will be based on the assumption that only scores of the first type, obtained by subtracting initial from final score, are to be considered valid measures of improvement. It is supposed, further, that irrespective of whether the original scores consist of amounts done in a given time, times re-

quired for a unit amount, or errors, they will be transformed, before subtracting initial from final score, so that the magnitude of the score increases with what by common sense is improvement in performance (23, pp. 10 and 14). Any of several methods are ordinarily satisfactory for such transformation. One method is absolute scaling, based on the percentages of a population exceeding each score (17; 19, p. 275). By this method any differences between the gains shown in amount-done scores and time-per-unit scores are caused to vanish. Instead of error scores, consisting simply of the number of errors, one may use the decrease in number of errors. Improvement would then be shown as a positive gain and not as a negative one.

The properties or characteristics of an individual which determine how much he will improve with any specified practice, under fixed environmental conditions, constitute his 'ability' to improve. In the case of those performances which consist of accomplishing a task (and only such performances are here considered), individuals ordinarily differ widely with respect to this ability so that differences therein are among the most important of all individual differences. The term ability, however, seldom refers to anything directly observable. It is very doubtful whether any abilities can be defined or identified except in terms of task-accomplishment. For example, we refer to the ability to add numbers correctly, to solve a puzzle, or, as in the case here to be particularly considered, to improve with practice in some specified performance. Similarly, by the ability known as intelligence we mean the ability to pass tests (providing the tests meet certain criteria) which have been arranged in order of difficulty. Likewise, the so-called group factors yielded by factor analysis, which are conceived as abilities, can be defined only by reference to the correlated performances. Thus by visual-spatial ability is meant merely a characteristic or set of characteristics which supposedly, to a considerable degree, determine an individual's scores in all of a certain set of performances, such as required by the tests known as punched holes, paper form-board, block-counting, etc. The definition of the ability lies in the description of these tasks.

It may be possible some day to describe some of man's innumerable abilities in neurological or biochemical terms. In other cases, it may be possible to identify an ability with a method of operation, that is, a particular pattern of processes constituting a way of going about a task. For the most part, however, it seems probable that our information concerning the nature of any ability will, for a long time to come, consist in a set of correlations between the ability and a list of other variables. This list of correlated variables may include other abilities, certain kinds of introspective reports, and certain kinds or amounts of past training. In view of this very incipient stage of our knowledge concerning the nature of abilities, it should not be surprising that what are here put forth as important facts concerning learning ability deal chiefly with correlations.

Providing the term learning ability is used in the sense more or less definitely outlined above, the following generalizations seem to hold with very little qualification:

1. The ability to learn cannot be identified with the ability known as intelligence.
2. Individuals possess no such thing as a unitary general learning ability.
3. Improvement with practice correlates importantly with group-fac-

tors, that is, relatively narrow abilities, and also with specific factors.

4. Even the group-factors involved in learning are not unique to learning, but consist of abilities which can be measured by tests given but once.

Since the term intelligence is here used in its customary sense, it means a single variable. Obviously, then, the first of the above statements follows as a corollary from the second; for the statement that there is no unitary general ability could not be true if intelligence, a single variable, could be identified with learning ability in all cases of learning. On the other hand, the truth of the second statement does not follow from the first, since by a general learning factor may be understood simply a variable which *in part* determines improvement with practice in all cases. Such a general factor could exist even though it were not in any particular case of learning identical with the total ability to improve. The first statement also is a corollary of the third, postulating the importance of group and specific factors; for if intelligence or any other single factor could be *identified* with learning ability, no room would be left for any other factors. On the other hand, the truth of the second proposition, postulating the absence of any general learning ability, would guarantee an important role for either group or specific factors, or both. The fourth proposition is independent of the other three. While all the above propositions are at some variance with views which seem to be rather widely held, it is believed they are substantiated by the experimental evidence. A review of this evidence will incidentally serve to elucidate their meaning. Finally, an attempt will be made to offer an explanatory interpretation of the propositions in the belief that thereby they will become more serviceable.

II. PREVIOUS VIEWS

Statements identifying learning ability with intelligence are found so frequently that a careless reader might form the opinion that such identification is beyond dispute and the evidence in support of it is so well known that there is no need to present it. Examples of such statements are the following: "intelligence is the ability to learn" (1); "an individual possesses intelligence insofar as he can learn" (3); and, "intelligence means capacity for learning" (5, p. 62). Officers are told that the army's General Classification Test is designed to determine a soldier's general intelligence *and* ability to learn. McGeoch, in his authoritative book on human learning, writes that "it is a reasonable hypothesis that reliable and extensive samples of both intelligence tests and learning experiments would be found to measure the same variables almost entirely" (13, p. 251). Even more definite is a statement by an investigator of learning that "Intelligence tests are constructed to indicate how much the child has *acquired* from general contact with his environment. They are tests of improvability" (4). Quotations from a number of prominent writers could be given indicating that the view is widely held that "The measurement of the actual progress of learning would furnish the best test of intelligence" (7).

No one, so far as the writer is aware, has marshalled the evidence in support of the view expressed in the above quotations. The implied arguments, however, seem to be variations or amplifications of the two following lines of thought.

1. The tests used to measure intelligence call for learned performances.

How well a person does on these tests, therefore, measures how well he can learn, and consequently his learning ability. This argument clearly deals with what has been termed past learning. It ignores future learning. At best it could only be argued that intelligence was identical with the ability to learn the particular tasks used in the measurement of intelligence. Even for such tests, however, it begs the issue by confusing achievement in these tasks, which is what actually determines the intelligence score, with the ability to gain with practice. As will be proved, later, the fact that intelligence correlates highly with score in a test signifies practically nothing concerning the degree to which intelligence correlates with amount of gain with practice in the test. Moreover, it is generally conceded that achievement in a test, even though learning may be required for any achievement at all, is due partly to the maturation of hereditary potentialities, which, though intimately related to learning, may be conceived in abstraction from learning. If this abstraction be made, clearly achievement, even in intelligence tests, could not be alleged to be due solely to learning.

2. Correlations between intelligence and scores in school achievement tests are ordinarily high. The argument is then that since school achievement tests measure learned performances and correlate highly with intelligence tests, learning ability is identical with intelligence. This argument involves the same fallacy as the preceding one when it attributes achievement solely to learning. Further, however, the correlation between achievement scores and intelligence scores is far from perfect. According to St. John (14), who averaged the results of 120 investigations, it is only .56. If the true correlation were assumed to be as high as .707 (the square root of .5), and it were further assumed that all of intelligence was a part of the total cause of goodness of achievement, we would be left with the conclusion that factors totally uncorrelated with intelligence have as much to do in determining achievement as has intelligence (9).

One cannot help wondering in connection with these arguments whether writers who identify intelligence with learning ability include under learning ability the ability to learn tasks which have little or no correlation with intelligence. Certain motor skills, in particular, show no appreciable correlation with intelligence either before or after practice, and yet show marked improvement with practice. It is hard to understand why the well known existence of such tests should not have produced more caution on the part of those who identify intelligence with learning. Certainly, when a performance shows no considerable relation to intelligence at any stage of practice, some very definite evidence is required before we can assume that improvement with practice therein is identical with intelligence. But no such evidence exists.

III. Evidence from Practice in Psychological Tests

Turning now to studies where there has been actual measurement of improvement in performance, we find that the results are opposed to the identification of intelligence and learning ability. Data obtained from well controlled laboratory experiments on practice are on the whole in good agreement with those obtained in schools upon the effect of such study and other exercises as constitute the educational program of the class-room. The results of typical studies will be briefly reviewed. In illustration of the results obtained from laboratory experiments, may be cited a study in which fifty-six subjects were given practice for thirty-nine days in

each of seven tests (18). The performances practiced represented a wide variety of activities and were the following: horizontal adding, substitution, reproduction of spot-patterns, rearranging letters to make words, cancelling letters with complex instructions, estimating lengths, and speed of making 'gates' (making four horizontal lines and one diagonal one in each square of a page divided into one thousand squares). The improvement score used was the difference between final raw score and initial raw score. Both initial and final scores had exceptionally high reliability (in the neighborhood of .94); and the practice was long enough so that for the most part the individual learning curves showed a pronounced flattening out towards the end of the practice. Only one of the gain scores showed a significant correlation with intelligence, as measured either by the Thorndike CAVD or the Otis Advanced Examination. With the Thorndike intelligence tests, three of the correlations shown by the gains were negative, and the average for the seven gains was .108. With the Otis tests, four of the correlations with gains were negative, and the average was .075. Clearly, in these cases, intelligence, far from being identical with the amount of improvement shown by practice, had practically nothing to do with the matter.

As regards the presence to any important degree of any other common factor sufficient negative evidence is afforded by the fact that the average of the twenty-one intercorrelation of the seven improvement scores was only .098. Six of them were negative and only one of the positive correlations was statistically significant. As might be expected, the results of a factor analysis clearly demonstrated the absence of any factor common to all seven of the improvement scores. Although no general factor common to all or even a majority of the gain scores was discovered, several factors common to two or three gain scores were quite conspicuous. One of these, for example, a factor similar to Thurstone's factor P, showed correlations with improvement in substitution of .50 and with improvement in cancellation of .91.

In a second experiment, eighty-two subjects were given practice for sixty-six periods in four tests. This experiment showed that the gains with practice could not be identified with intelligence, but were instead correlated with group and specific factors (21). It is true that a factor common to the four gains was found, but, since the correlations analyzed in this study included numerous 'end-tests,' it was possible to break up the general factor into several group factors. Of these the most significant ones were analagous to Thurstone's factors N and P (numerical and perceptual factors). The second study thus confirmed the first one in showing little relation between intelligence and what has been termed future learning, and considerable dependence upon group factors of the sort that may be defined in terms of test given but once without reference to learning. Both studies likewise agree in showing a marked relation between specific factors and the observed gains, indicating that gain in any one performance is related to factors not concerned with gains in any of the other performances.

In an investigation just completed and as yet unpublished (15), Simrall gave practice to ninety-five high school students in two tests. The group included students from all four years of the University of Illinois High School. One of the practice tests was a jumbled words maze test and the other a backward writing test. Practice was sufficient to produce a gain in mean score from 157 to 290 in the first test and from 89 to

216 in the second one. Both tests showed rather marked correlations with intelligence (determined by averaging score on forms A and B of the Otis Advanced Intelligence Examination). Mental age correlated .60 with initial score and .61 with final score in the case of the backward writing test, and .59 with initial score and .49 with final score in the case of the jumbled words maze test. Correlation of mental age with the gain due to practice, however, was only .277 in the case of the backward writing test and − .079 in the case of the jumbled words maze test. The correlation between gain in one test with gain in the other was only .175. Sizable correlations between gains and group factors were found. The group factors, however, were not unique to these gains, but were identifiable by certain of Thurstone's tests known to involve these factors. Otherwise, in fact, they would not have been detected.

Conclusions similar to those reached in the studies just described were also drawn by Heese (10), from the correlations of gain scores of fifty subjects to whom he gave ten practice periods in each of six tests. He used as measures of the improvement of any subject the change in that subject's standard scores. A standard score for any trial was the deviation from the group mean divided by the standard deviation for that particular trial. Changes calculated from such scores, like changes in rank, may not correlate very highly with changes in raw score, so that the difference in the two kinds of measures is more than simply one in the choice of units. The fifteen intercorrelations between the gain scores were all positive, but nine of them were insignificant. From this fact, and also from the results of a factor analysis of the intercorrelations, Heese concludes that there was no general factor in the gains. He found three common factors, each of which appeared to show significant relation to the gains in two or three of the six tests. These factors were tentatively interpreted as speed of movement, memory, and perception, factors certainly not unique to learning scores. His estimates of the reliability of his gain scores show an average of .78 while the average percentage of variance accounted for by the three factors taken together was .47. Calculations which may be made from these figures [1] reveal that on the average the gains were due more to factors specific to the single gain scores than to the sum of the common factors. Heese's results should perhaps be regarded as extending rather than confirming those obtained in the experiments previously described. They extend them in the sense that they show that the same conclusions hold for changes in standard score as for changes in raw score.

One other study to determine the presence or absence of a general factor in learning scores was conducted by deWeerdt (4), who used as subjects forty-five fifth grade pupils averaging ten years in age. The gain scores were correlated with the measures of intelligence afforded by the Illinois Examination, Form I. Only one of these correlations was significant, a correlation of .56 with gain in speed of reading. The remaining six correlations between gain and intelligence varied from .28 to − .07. The average of the seven correlations was only .19.

The results obtained by Husband should probably not be included in this review, as the scores which he corre-

[1] The relation between variance due to group and to specific factors is expressed by the equation $\sigma^2_s = \sqrt{r_{11} - h^2}$, in which σ^2_s is the variance in standard score attributable to the specific factor, r_{11} is the reliability of the score, and h^2 is the variance in standard score attributable to the sum of all the group factors revealed by the factor analysis of the particular variables included in the study.

lated were apparently the total scores, consisting ordinarily of scores made in a number of successive trials, on what are frequently termed learning tests, such as maze learning, mirror drawing, etc. He made no calculation of gain as distinct from raw achievement. Nevertheless, it is interesting to note that he found exceedingly low correlations, in two studies, between total scores on the learning tests and also between learning scores and intelligence. In discussing his results, he writes as follows: "We realize there is no special reason for assuming that there should be a general learning ability, beyond its commonly being mentioned tacitly as if it were such. We also tend to assume that because learning is a fairly complex function it must be fairly well correlated with general intelligence, which again would suggest that it is a fairly unitary ability" (11). He finds, however, that learning scores and intelligence are virtually unrelated. "Correlations with intelligence are so weak as to be practically negligible." "The whole trend of evidence is such as to suggest that we must speak of learning abilities (plural) and not of learning ability (singular) as if it were a general ability" (12).

IV. Evidence from School Learning

Exercises properly included under the concept of practice constitute to a large extent the courses of study in schools. Learning to read means largely changes in behavior due to practice, some of it corrected or controlled practice, some of it uncontrolled, some of it in school and some of it out of school. The same thing may be said of learning to spell, of learning to make arithmetical computations, to think out the solution of arithmetical problems, to use English, to punctuate, to employ words grammatically, etc. It is, therefore, of much interest to see whether the same relations hold between gains in school subjects as between gains revealed by studies of practice in psychological tests. Several recently conducted investigations are available. They show the same lack of interrelation between gains in school subject matters or skills, and the same lack of correlation with intelligence, as do the gains observed in the laboratory experiments.

Two studies of gains in school subjects over periods of from one to three years in the grammar school grades have been recently published (24). One study dealt with the gains of pupils in Litchfield, a small town in Illinois, and the other with those shown by the pupils in one of the schools of Providence, Rhode Island. Objective achievement scores were obtained for six school subjects, namely, reading, vocabulary, arithmetic fundamentals, arithmetic problems, English, and spelling. These scores were the part-scores yielded in one case by the Metropolitan Achievement Battery and in the other case by the Stanford Achievement Tests. In all cases the intercorrelations of the gains made in the various school subjects were very low. In the Illinois school, the average of forty-five such gain intercorrelations was only .12, in striking contrast to the average intercorrelation of .60 between the test scores on a single date. In the Rhode Island school, the average of forty-five intercorrelations between gain scores was .13.

The general conclusions indicated by the two studies were found to be as follows: no general gain factor; lack of significant relation between change in score and IQ in grades beyond the fifth; a group factor common to the two arithmetical gains (in problems and in fundamentals); a group factor common to gains in reading and in other verbal tests; and indications of the importance of the specific factors.

Likewise, in the case of college stu-

dents, there appears to be no correlation between gain in achievement test score and intelligence. Drake studied this matter in the case of progress in biology and history. He found, for example, a correlation of $-.14$ between intelligence as measured by the American Council Psychological Examination and gain in knowledge of biology as measured by two forms of the Coöperative Test Service Biology Test. As a measure of gain, he used the difference between initial and final standard scores. He believed that the explanation of this result necessitates the postulation of a growth factor not related to intelligence, to which he applies the term 'Iota Function' (6).

Similar results were obtained by Dysinger and Gregory in a careful study (8) of the gain in knowledge from a course in elementary psychology. An objective examination was given to students in four sections, at the beginning and the end of the course. The difference between the scores at final and initial testing was taken as indicative of the improvement during the semester. This gross improvement score showed a correlation of $-.064$ with scores made on the Army Alpha Group Intelligence Test (Nebraska Revision). This result was confirmed by a similar study carried out with students taking a semester course in introductory psychology at the University of Illinois (2). The correlation between gain in objective test score and intelligence as measured by the Otis Gamma test was found to be $-.039$, surprisingly close to that obtained in the earlier study.

Dysinger and Gregory noted, further, that those students who obtained high scores in the first test could not obtain a high gross improvement score, because, owing to the limits of the test, the number of questions on which they could improve obviously would be fewer than the number upon which students making a low initial score could improve. Consequently, a second measure of improvement was calculated. This measure consisted of the ratio between the gross improvement and the maximum possible improvement (total number of test questions minus the number of correct replies on the first test). These ratios also failed to show any significant relationship with the intelligence test score. The obtained correlation of .12 indicates of course that the relative improvement achieved during the semester as measured by these ratios was almost entirely independent of the level of ability as measured by the Army Alpha score.

V. Evidence from the Rate of Learning

The numerous and extensive studies which have been reviewed are without exception in support of the four propositions enunciated earlier in this paper. It may be objected that the validity of the obtained results may be limited by the fact that it has always been impossible to conduct an experiment over such a period that the entire learning or practice of the subjects has been covered. The subjects do not start at an absolute zero and do not ordinarily reach their theoretical upper limit. An experiment covering such a range is probably not feasible. In the case of school subjects it would obviously have to extend over a considerable number of years. At what age of the subjects it would have to begin is unknown, because the zero of performance is a matter so shrouded in doubt as to be definitely a theoretical topic.

As a result of the limitations which it appears necessary to place upon practice experiments, it follows that a gain experimentally measured is only a part of some larger gain, an unknown portion of which may have occurred before the experiment began and another

unknown portion of which (ordinarily much smaller) may occur after the experiment has ended. Until we are able to conduct experiments extending over the lifetime of individuals, with their whole lives under experimental control, the best hope of estimating the influence of the limitations on practice experiments seems to be in fitting individual practice curves to the experimental data. Providing the test scores are highly reliable and the experiment extends over a considerable period, a practice curve may be fitted to the data of each subject with a fair degree of reliability. Now once this laborious curve-fitting job has been done, it becomes possible to calculate the values of certain measures of learning independently of the portion of a subject's practice curve which is covered by the experiment. One such value, perhaps the best conceivable measure of learning, is the amount of gain in unit time at that point in the subject's learning curve where his gain is most rapid. Mathematically stated, this measure is one proportional to the first derivative of the learning curve at the point in the learning curve where the curve is rising most rapidly. One curve, which fits most practice data quite well, enables one to calculate this value rather easily. It is a curve which shows this most rapid rise at the point where the curve has attained a height equal to one half the indicated upper limit. This point may occur in a given subject at any stage of the observed practice or even at a point in the curve preceding that portion which represents observed practice. It does not depend, therefore, upon what point in his practice curve the subject happened to be on the day the practice experiment began. To permit measures of rate of learning at this point, four hundred and ninety-six curves were fitted to the practice data obtained from fifty-six subjects given practice in each of three performances and from eighty-two subjects given practice in each of four performances (22). From each fitted curve was calculated the gain per unit time at the point where the curve was rising most steeply. These rates of gain were then correlated with each other. In general, the intercorrelations of these measures of rate of gain were insignificant. So also were their correlations with intelligence. The intercorrelations between rates of gain averaged .146 and the correlations between rate of gain and intelligence averaged .085.

Obviously the results of experiments made in either the laboratory or the classroom, in which improvement with practice has been actually measured, are all in contradiction with the assumption, based on general considerations, that the ability to learn in the sense of the ability to improve with practice is identical with intelligence. Far from identity, the experimental studies show that as a rule the correlation between intelligence and gain, while usually positive, is so small as to possess little significance. Only occasionally is it large enough to be of moderate importance; often it is very near zero. Moreover, in no experiment with human subjects have gains in a wide assortment of performances been found to have any general factor. On the other hand, factors of considerable importance have ordinarily been discovered in the gains shown in a limited group of performances. Among these group factors, however, there is none which is peculiar to learning. They are of the sort which can be measured by ordinary testing techniques consisting of giving tests but once, and include such factors as memory, visual-spatial ability, speed, and perceptual abilities. There has also been obtained abundant evidence of the importance of factors which are uncorrelated with

the group factors and which are uncorrelated with each other.

VI. INTERPRETATION

Some difficulty in accepting the conclusions stated above may be experienced in certain quarters on account of a tendency to confuse achievement with gain. This confusion has prevented a clear understanding of how it is possible for intelligence at all stages of practice to have a high correlation with achievement and yet to have little correlation with gain due to practice. Perhaps the best way to clear up this confusion is to consider gains in the light of two widely accepted formulae. One of these, which has been tacitly assumed in much of the above discussion, is the factorial equation as written from the two-factor viewpoint. If we have a few tests containing no common factor but G, which for the present purpose will be considered a substitute for the concept of intelligence, then the score in any test correlating with G may be divided into two summative parts, the first of which is the amount due to G, and the second the amount due to s, the factor or factors specific to each test. The equation for score at any stage of practice may then be written as follows: $y = c_1G + c_2s$, in which G is the general factor, s the remaining determining factors, y is score, and c_1 and c_2 are coefficients. For the sake of simplicity, all measures are assumed to be free of error. If we now represent the contribution of G to initial score by g_1 and to final score by g_2, improvement with practice may be expressed in terms of the factorial equation. The improvement equals $g_2 + s_2 - g_1 - s_1$, in which s_1 and s_2 represent, respectively, the contribution of the specific factor to the initial and the final practice scores. That the specific factor changes with practice has been experimentally demonstrated (21, pp. 67–68). On the other hand, there is no reason to believe that s_1 and s_2 would not be correlated. We may now write the correlation between G and gain, that is, $r_{(G)(g_2+s_2-g_1-s_1)}$, by use of a second formula, that for the correlation of sums or differences (16). Since the correlation between G and either s_1 or s_2 is zero (by factor theory), the correlation between G and gain is equal to $(r_{Gg_2}\sigma_{g_2} - r_{Gg_1}\sigma_{g_1})$ divided by

$$(\sigma^2_{g_2} + \sigma^2_{s_2} + \sigma^2_{g_1} + \sigma^2_{s_1} - 2r_{g_1g_2}\sigma_{g_1}\sigma_{g_2} - 2r_{s_1s_2}\sigma_{s_1}\sigma_{s_2})^{\frac{1}{2}}.$$

It might seem that r_{Gg_2} and r_{Gg_1} should in the case of errorless data both be unity. This might seem to be true since both g_1 and g_2 equal G times a constant. But these constants are averages. They could and presumably do vary from individual to individual. It seems plausible, therefore, to assume a correlation of less than unity between G and g_1 and g_2. Moreover, it is within reason to imagine, on account of the increase in variability which usually occurs with practice (which may be construed to indicate greater variability in manner or performance after practice than initially), that r_{Gg_2} may often, and perhaps usually, be less than r_{Gg_1}. It is likewise reasonable to suppose that σ_{s_2} is ordinarily greater than σ_{s_1}, and that in the case of experiments running over a long time σ_{g_2} is greater than σ_{g_1}. Finally it is reasonable to suppose a substantial correlation between s_2 and s_1. The correlation between g_2 and g_1 will be the product of their correlations with G.

Along the lines of the above assumptions, suppose, for the sake of illustration, we attribute specific quantitative values to all the variables involved in writing the correlation with G in terms of both factor theory and the formula for sums. Let first the following values be supposed: $\sigma_{g_1} = 1$; $\sigma_{g_2} = 1.2$; $\sigma_{s_1} = 1.0$; $\sigma_{s_2} = 1.5$; $r_{s_1s_2} = .50$; $r_{Gg_1} = .9$; $r_{Gg_2} = .8$; $r_{g_1g_2} = .72$. The resulting cor-

relation between G and gain is .04. In spite of this low correlation between G and gain, the correlation between G and initial score is .64 and between G and final score, .50. Again, suppose all the above assumptions are kept the same except that r_{Gg_2} is assumed to be .85 (in which case $r_{g_1g_2}$ becomes .765) and σ_{s_2} is assumed to be 1.0. Then the correlation between G and initial score, as before, is .64; the correlation between G and final score is .65; and the correlation between G and gain is only .09. This last illustration shows how in spite of high and undiminishing correlation between G and score, at all stages of practice, it is quite possible for the correlation between G and gain to be negligible. These illustrations should abolish the argument that when intelligence correlates highly with achievement, ability to improve in achievement with practice must be largely a matter of intelligence. Incidentally, the first set of assumptions, according to which correlation of score with G decreases from an initial correlation of .64 to one of .50 at the end of practice, serve to explain why it is a common thing (20) for the correlation of a test with intelligence to decrease with practice.

The illustrations given also show how it is possible that two tasks, both of which correlate with intelligence, display very little in the way of a common gain factor. For example, two tests each correlating .64 with intelligence, would according to either of the two sets of values assumed above, show a correlation of about .41 with each other at the beginning of practice. The correlation between gain in one and gain in the other, however, would be practically zero in the case of either set of assumptions.

In this discussion, the presence of group factors has been ignored. To assume one or more group factors in addition to the general and specific factors considerably complicates the task of writing the formula for the correlation of G and gain in terms of factors. The outcome of such formulae will, of course, depend upon the numerical values assumed. The postulation of group factors, however, does not result in any particular difficulty in showing how high correlation between G and achievement is compatible with little correlation between G and gain, and little correlation between gains in different tests. Moreover, if one is considering only a small group of tests, and these, while all correlating with intelligence, happen to be of sufficient variety that no two of them have the same group factor, all the group factors vanish, that is, become indistinguishable from the specific factors.

In general, the statistically discovered high degree of specificity of practice gains may be interpreted as meaning that each particular activity can best be performed by methods which to an important degree are peculiar to that activity. Improvement with practice depends very largely upon the adoption at repeated trials, to a degree varying with the individual, of these particular methods.

REFERENCES

1. BUCKINGHAM, B. R. Intelligence and its measurement: a symposium. *J. educ. Psychol.*, 1921, 12, 271–275.
2. CARLSON, H. B., FISCHER, R. P., & YOUNG, P. T. Improvement in elementary psychology as related to intelligence. *Psychol. Bull.*, 1945, 42, 27–34.
3. CARROL, R. P. What is intelligence? *Sch. & Soc.*, 1928, 28, 792 (quotation from S. S. Colvin).
4. DEWEERDT, E. H. A study of improvability of fifth grade school children in certain mental functions. *J. educ. Psychol.*, 1927, 18, 547–557.
5. ELLIS, R. S. *The psychology of individual differences.* New York: Longmans, Green and Co., 1929.
6. DRAKE, C. S. The Iota function. *J. educ. Res.*, 1940, 34, 190–198.

7. DEARBORN, W. F. The measurement of intelligence: a symposium. *J. educ. Psychol.*, 1921, **12**, 211.
8. DYSINGER, D. W., & GREGORY, W. S. A preliminary study of some factors related to student achievement and grades in the beginning course in psychology. *J. gen. Psychol.*, 1941, **24**, 195–209.
9. GARRETT, H. E. *Statistics in psychology and education.* New York: Longmans, Green and Co., 1926.
10. HEESE, K. W. A general factor in improvement with practice. *Psychometrika*, 1942, **7**, 213–223.
11. HUSBAND, R. C. Intercorrelations among learning abilities. *J. genet. Psychol.*, 1939, **55**, 353–364.
12. ——. Intercorrelations among learning abilities: IV. Effect of age and spread of intelligence upon relationships. *J. genet. Psychol.*, 1941, **58**, 431–434.
13. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green and Co., 1942.
14. ST. JOHN, C. W. *Educational achievement in relation to intelligence.* Cambridge, 1930.
15. SIMRALL, D. V. The effects of practice on the factorial equations for perceptual and visual-spatial tests. Ph.D. Thesis, University of Illinois Library, Urbana, 1946.
16. SPEARMAN, C. Correlations of sums or differences. *Brit. J. Psychol.*, 1913, **5**, 417–426.
17. WOODROW, H. The scaling of practice data. *Psychometrika*, 1937, **2**, 237–247.
18. ——. The relation between abilities and improvement with practice. *J. educ. Psychol.*, 1938, **29**, 215–230.
19. ——. The effect of practice on groups of different initial ability. *J. educ. Psychol.*, 1938, **29**, 268–278.
20. ——. The effect of practice on test intercorrelations. *J. educ. Psychol.*, 1938, **29**, 561–572.
21. ——. Factors in improvement with practice. *J. Psychol.*, 1939, **7**, 67–68.
22. ——. Interrelations of measures of learning. *J. Psychol.*, 1940, **10**, 49–73.
23. ——. The problem of general quantitative laws in psychology. *Psychol. Bull.*, 1942, **39**, 1–27.
24. ——. Intelligence and improvement in school subjects. *J. educ. Psychol.*, 1945, **36**, 155–166.

A CRITIQUE OF 'NEUTRAL' ITEMS IN ATTITUDE SCALES CONSTRUCTED BY THE METHOD OF EQUAL APPEARING INTERVALS

BY ALLEN L. EDWARDS

University of Washington

I. Introduction

By any standard of judgment *The Measurement of Attitude* by Thurstone and Chave would deserve a significant position in the annals of social psychology. Since the publication of this monograph, in 1929, hundreds of articles which have appeared in psychological, educational, and sociological journals have either made use of scales constructed by the Thurstone and Chave technique or have had as their major concern the construction of such scales.

The method developed and advocated by Thurstone and Chave for the measurement of attitude is the psychophysical method of equal appearing intervals. In application the method requires that a judging group sort statements of opinion along a continuum ranging from 'favorable,' through 'neutral,' to 'unfavorable.' The scale value of each statement is then determined by locating the point on the continuum below which and above which 50 per cent of the judges place it. Ambiguity of an item is measured by the spread of the judges' ratings along the continuum—quantitatively in terms of '*Q*.' The 'relevance' of an item can also be determined by a criterion developed by Thurstone and Chave which is based upon an 'index of similarity.'

With the publication of Likert's monograph, '*A technique for the measurement of attitude*,' in 1932, the preeminence of the method of equal appearing intervals in the construction of attitude scales was seriously challenged.[1] In this monograph, Likert described a new procedure which was later to become known as the method of summated ratings.[2] This method, instead of using a judging procedure to sort items, made use of a voting procedure in which subjects are requested to check 'Strongly Agree,' 'Agree,' 'Undecided,' 'Disagree,' or 'Strongly Disagree,' in response to each item. By finding the per cent responding with each alternative and by referring to the tables of the normal curve, the alternatives may then be assigned scale positions in terms of sigma values. Since Likert found, however, that the correlation between the sigma method of scoring items and arbitrarily assigning weights of 1, 2, 3, 4, and 5, to the various alternatives was quite high, the latter method is generally used. From the initial large group of items, a maximum of about 25 are selected by means of a criterion of internal consistency for inclusion in the scale.

It was claimed that this method of scale construction possessed certain advantages over the method of equal appearing intervals: "it avoids the difficulties encountered when using a judging group to construct scales" as required by the Thurstone procedure (12,

[1] Three subsequent publications, Hall's study of 'Attitudes and unemployment,' the Rundquist and Sletto study of *Personality in the Depression*, and *Public Opinion and the Individual* by Murphy and Likert, were influential in directing attention to the Likert method of scale construction.

[2] It was Bird (1, p. 159), not Likert, who first used this term to designate the technique.

p. 42); it is much easier to apply (12, p. 43); it yields reliabilities as high as those obtained with the Thurstone scales, yet uses fewer items (12, pp. 42–43); it gives results which are comparable to those obtained by the Thurstone method (12, p. 26).[3]

If these claims are valid, then surely there is no practical reason why the method of summated ratings could not and should not be used in place of the method of equal appearing intervals in the construction of attitude scales. And it would seem that there is no practical reason why a judging group, in the Thurstone sense, is a necessary step in the construction of attitude scales.

II. Ferguson's Criticism of the Likert Technique

The advantages claimed for the method of summated ratings and the implication concerning the judging group, however, have been questioned by Ferguson (4) on theoretical rather than practical grounds.[4] By re-scaling with the Thurstone technique the items from a series of scales originally constructed by the Likert method (17), Ferguson found that the Likert items failed to spread evenly over the scale continuum posited by Thurstone and Chave. Very few items fell in the middle section, the so-called 'neutral' area of the continuum, the majority piling up at the 'favorable' and 'unfavorable' ends. Ferguson believed, therefore, that he had demonstrated that the method of summated ratings did not "obviate the need for a judging group" (4, p. 57).[5]

What Ferguson has demonstrated, I believe, is nothing more or less than what he found: that items included in a scale constructed by the Likert method will tend to be items which, if scaled by the method of equal appearing intervals, have scale values at the extremes of the Thurstone continuum. Whether this finding has any bearing upon 'the need' for a judging group depends upon whatever value 'neutral' items have in an attitude scale and in the measurement of attitude—a problem we shall take up later.

The explanation for Ferguson's finding that 'neutral' items tend to be scarce in Likert scales has its basis, I believe, in the technique of item selection used in the construction of scales by the method of summated ratings. Criterion groups, in terms of low total scores and high total scores, are selected and the responses of these two groups to each item are compared. Items which fail to show a significant difference between the two groups are eliminated. In terms of the criterion of internal consistency, such items would be designated as non-differentiating items.

III. 'Neutral' Items as Non-Differentiating Items

It is my contention that Thurstone's 'neutral' items are precisely of this nature. They are non-differentiating

[3] The first three advantages are explicitly claimed, the fourth is implied. The specific quotation on the page referred to is: "It seems legitimate to inquire whether it (Thurstone's method) does its work better than the simpler scales which may be employed . . ." (12, p. 26).

[4] See also the discussion by Bird (1, pp. 160–164).

[5] Miss Claire Kenney and I have reviewed the evidence bearing upon the controversy which has centered around the relative merits of the two methods of scale construction in a recent article (2). It was our contention that Ferguson's conclusion concerning the need for a judging group was based upon an inadequate test of the hypothesis involved. We demonstrated, at least in the one instance we were concerned with, that it was possible to take the same set of original items, to construct scales independently by the Thurstone and Likert techniques with this set of items, to administer both scales to a new group of subjects, and to obtain substantial correlations between scores on the two independently constructed scales.

items. Consequently, they would not be expected to appear in any great number in scales constructed by the method of summated ratings.

If the hypothesis is essentially correct, then items found to have low and high scale values, when sorted by the Thurstone technique, should have high mean difference values in terms of Likert's criterion groups. Items with Thurstone scale values in the middle or 'neutral' section of the continuum should tend to have low mean difference values when tested by the criterion of internal consistency.

To see whether this relationship existed or not, I analyzed the scale and mean difference values for a set of 129 items which Thurstone and Chave had used in the construction of their scale to measure attitude toward the church.[6] These items had been treated by both the Thurstone and Likert techniques in the process of preparing the article referred to earlier (2) and, consequently, Thurstone scale values and Likert mean difference values were available for each item.

A bivariate distribution was plotted, the Thurstone scale values being represented by the abscissal scale and the mean difference values by the scale of ordinates. The resulting distribution was U-shaped with low and high scale values, as determined by the method of equal appearing intervals, associated with high mean difference values. Items with Thurstone scale values toward the center of the continuum were associated with low mean difference values.

The null hypothesis, that the true correlation between the two variables is zero, was tested by means of *epsilon-square*. In the present analysis, N, the number of items, is equal to 129; k, the number of columns in the correlation table, is equal to 15, and ϵ^2 is equal to .2432. According to the tables of ϵ^2 published by Peters and Van Voorhis (13), if the true correlation is zero, then a value of ϵ^2 of .2432 is highly significant when $N - k = 114$ and $k - 1 = 14$.

A more positive statement concerning the relationship between the mean difference values and scale values of the items can be made in terms of the square root of ϵ^2, which is an unbiased correlation ratio (13, pp. 319–325). This is equal to .49 and indicates the strength of the relationship between the two variables.

Further evidence bearing upon the contention that the Likert method of item selection would tend to eliminate items with scale values in the 'neutral' section of the Thurstone continuum is available if we examine the scale values of the 25 most differentiating items—the (maximum) number ordinarily included in a Likert scale. The 129 items were sorted by the Thurstone technique on a continuum ranging from 1 to 9. Eight of the 25 most differentiating items have scale values under 1.1 and 12 have scale values 7.8 or greater. The remaining 5 items are distributed as follows: 2 between 5.4 and 5.9; 1 between 6.0 and 6.5; 1 between 6.6 and 7.1; and 1 between 7.2 and 7.7. In other words, only 2 of the 25 most differentiating items have scale values anywhere near the 'neutral' area of the continuum.

Thus the evidence would seem to indicate that the Thurstone 'neutral' items are non-differentiating items and for this reason would not be found to be represented heavily in scales constructed by the method of summated ratings. In view of this finding, we might well ask what function these items serve in an attitude scale? What do they contribute to the measurement

[6] There were 130 items in the original Thurstone and Chave list. One of the items was inadvertently omitted when the set was mimeographed for sorting purposes.

of attitude? Why are these items judged to be 'neutral,' *i.e.*, sorted into the middle categories of the Thurstone continuum by the judging group, in the first instance?

IV. JUDGMENT OF ITEMS

If we examine the use which is made of the concept of 'neutrality' by Thurstone and Chave, we find it is not defined very precisely. At one point the statement is made, referring to the continuum, that:

> "The base line represents ideally the whole range of attitudes from those at one end who are most strongly in favor of the issue to those at the other end of the scale who are as strongly against it. Somewhere between the two extremes on the base line will be a neutral zone representing indifferent attitudes on the issue in question" (19, p. xi).

'Neutrality,' in this instance, would seem to be synonymous with 'indifference,' and this interpretation is supported by a later statement concerning a scale ranging from militarism to pacifism "with a neutral or indifference zone between" these two extremes (19, p. 14).

But elsewhere, in a discussion of a prohibition scale, we find that this "neutral zone" would "include neutral *and* indifferent attitudes" (19, p. 11).[7] And, at another place in connection with a discussion of a minor study involving self-ratings on the continuum, we find the statement: "It may be also that some of our subjects failed to understand the scale and interpreted 'neutral' to mean complete or active indifference to the church" (19, p. 80).

Now, in view of these statements, I am not at all sure whether Thurstone and Chave intend 'neutrality' to be synonymous with 'indifference' or not. If a distinction is to be made between the two terms, then what are the essential characteristics of 'neutrality'? Unfortunately, Thurstone and Chave provide neither us nor the judging group with the answer to this question.

The judges are working with various categories with only the end points and the center position anchored.[8] One extreme, the judges are told, "expresses strongest appreciation," the other, 'strongest depreciation.' The middle pile, they are told, "contains only neutral expressions" (19, p. 30). The judges may have little difficulty in interpreting the two extremes of this continuum, but what will they take the statement "contains only neutral expressions" to mean?

The evidence we have from our previous analysis is that the items sorted in the 'neutral' section of the scale tend to be non-differentiating. But items may be non-differentiating for a number of reasons. They may be non-differentiating because they are *ambiguous*, open to different interpretations, capable of being understood in different ways. They may be non-differentiating because they are *ambivalent*, express indecision. They may be non-differentiating because they are *irrelevant*, contribute nothing to the location of the individual on the scale, or are extraneous to the issue. They may be non-differentiating because they are *indifferent*, express lack of concern or apathy with respect to the issue.[9] Judges may, in other words, place an item in the middle categories because it is interpreted as expressing any one of these conditions. The middle cate-

[7] Italics not in the original quotation.

[8] It is necessary that the other categories remain undefined in order that "the intervals between successive piles should be apparently equal shifts of opinion as judged by the subject" (19, p. 30).

[9] And, of course, they may be non-differentiating because they are factual—or at least interpreted as statements of fact rather than as statements of opinion.

gories, on the Thurstone continuum, may be simply the 'catch-all' categories.

V. 'NEUTRAL' ITEMS AND THE CRITERION OF AMBIGUITY

All of the hypotheses raised concerning the reason why judges sort items into the middle categories of the Thurstone continuum may be tenable. Let us consider first the possibility that items with scale values in the center of the scale tend to be relatively ambiguous. Ambiguity of an item is determined by Thurstone and Chave in the following way:

"If the 300 subjects place a statement of opinion in widely different intervals on the subjective scale, the Q-value of the statement will be large and the statement will therefore be judged by this objective criterion to be ambiguous. It has widely different meanings along the attitude scale when it is read by different subjects" (19, p. 46).[10]

[10] Thurstone and Chave also expressed concern over the problem of the independence of Q and scale values.

"Since the Q-value is to be used for eliminating ambiguous statements it becomes a matter of some importance to know if the average Q-value changes for different parts of the scale. In order to make this relation apparent we have plotted the average Q-value against scale-values. . . . If all the statements throughout the whole range were of the same average ambiguity, the line of this diagram would be horizontal and that would be the ideal condition for our present purposes. It does not show any gross variation in ambiguity although the statements which are in the vicinity of 7 or 8 on the scale have noticeably higher ambiguity than the other statements" (19, pp. 44–45).

By re-plotting the paired Q and S values in a correlation table, a precise test of the independence of the two variables is possible. For $N - k = 118$ and $k - 1 = 11$, I obtained a value of ϵ^2 equal to .2002. This is highly significant and the hypothesis of independence must be rejected. The strength of the association is given by the unbiased correlation ratio, the square root of ϵ^2. This is equal to .45. But we must note that in this test the maximum Q values do not appear at precisely the center of the continuum but at scale values 7 to 8 on the 11 point continuum.

To test the hypothesis that scale and Q values are independent, the paired scale and Q values for the 129 items mentioned earlier were plotted in a bivariate distribution. The plotted values indicated a trend in accord with the hypothesis advanced. Lower Q values are found at both extremes of the scale and higher Q values toward the center of the scale.[11] The significance of the association was tested by means of ϵ^2. The value obtained was .7144, a highly significant value for $N - k = 114$ and $k - 1 = 14$, and the unbiased correlation ratio, of course, is equal to .85.

Additional tests of the hypothesis were made with data published by various individuals working under the direction of Remmers (15). Pairs of scale and Q values were available only for the items retained in the completed scales, a total of around 90 for the two forms of each scale. But since, as one investigator points out, in these Thurstone constructed scales "statements with high Q values were not used" (8, p. 22), it should be even more significant if the 'neutral' sections of the scales tend to have high Q values. This is because the data should include items with minimum Q values selected from all parts of the scale continuum. And if the items with minimum Q values have been selected for all positions of the continuum, then we should expect that the Q values of the 'neutral' items will not be in excess of those at the two extremes.

The results of six different tests of the hypothesis are summarized in Table I. In each instance a significant value

[11] H. E. Miller (11, p. 70) shows a scatter diagram of Q and scale values for 159 items which were sorted by judges. This scatter diagram is very similar in appearance to the one I obtained with the 129 items referred to above.

TABLE I

SUMMARY OF TESTS OF SIGNIFICANCE OF RELATIONSHIP BETWEEN SCALE AND *Q* VALUES OF ITEMS IN SCALES CONSTRUCTED BY THE METHOD OF EQUAL APPEARING INTERVALS

Investigator	Attitude Scale	*N*	*k*	ϵ^2	ϵ
Kelley (8)	Any Institution	90	11	.3083	.56
Grice (5)	Any Group	92	11	.3787	.62
Kellar (7)	Any Home-Making Activity	90	10	.6708	.82
Miller (11)	Any Occupation	90	11	.5423	.74
Silance & Remmers (18)	Any School Subject	90	11	.5081	.71
Miller (10)	Teaching	90	10	.1695	.41

of ϵ^2 was obtained.[12] Clearly, then, we have evidence to indicate that the scale and *Q* values of items sorted by the method of equal appearing intervals are not independent, that 'neutral' items are relatively more ambiguous than items at either extreme of the continuum.

VI. 'NEUTRAL' ITEMS AND THE CRITERION OF IRRELEVANCE

The criterion of irrelevance, based on the index of similarity, if applied in the construction of scales by the method of equal appearing intervals, would seem to rule out the possibility raised earlier that items in the 'neutral' section of the scale continuum may be irrelevant to the issue under consideration.

"This criterion is constructed as follows: Suppose that a statement of low ambiguity is properly scaled at the point 6. If a subject has an attitude which is also scaled properly at the point 6, then we should expect him to check that statement. Another subject who is scaled at the point 12 should be less likely to check that statement, and similarly there should be a low probability that a subject at the point zero will check the statement at 6 on the scale" (19, p. 47).

[12] Each distribution of plotted points, of course, showed some irregularities. But the trend, in each instance, was in accord with the hypothesis, the plotted points resembling an inverted U.

The index of similarity which quantifies the analysis described above is based upon three factors for any pair of statements. For the pair of statements *a* and *b*, for example, these factors are: n_a = the total number of subjects endorsing statement *a*; n_b = the total number endorsing statement *b*; and n_{ab} = the total number endorsing both *a* and *b*.

"If the two statements *a* and *b* are practically identical in the attitudes they reflect, then we should expect to find that those subjects who indorse statement *a* will also indorse statement *b*. This factor, n_{ab}, will therefore be in the numerator of the index of similarity. On the other hand, the statements vary considerably in intrinsic popularity even when they are scaled at identical points on the scale. The more popular a statement is, the larger will be the number of people who indorse it and any other statement. In order to reduce the index of popularity to the same basis of popularity for all statements, the number of subjects who indorse both statements is divided by the product of the number of total indorsements for each of the two statements so that the index of similarity becomes

$$\frac{n_{ab}}{n_a \cdot n_b}.$$

"If we tabulate the indices for statement *a* with each of all the other statements in turn, we shall have the common factor $1/n_a$ which may be disregarded since it is a constant" (19, pp. 47–48).

The index of similarity thus becomes n_{ab}/n_b. If all of those who endorse statement *a* also endorse statement *b*, then the index is equal to unity; whereas if none of those who endorse *a* endorse *b*, then the index is equal to zero. The criterion of irrelevance for statement *a* is simply the appearance of a diagram in which the various indexes of similarity for statement *a* are plotted against the scale values of the statements. If an item meets the test of this criterion, then the values of the index should be high for items with scale values near the scale value of the common statement, in this case statement *a*, and should be low for items with scale values distant from this point.[13]

On the assumption that the criterion of irrelevance had been applied to the scales constructed under Thurstone's editorship, I gave four different scales (14, 16, 20, and 22) to subjects at the University of Washington.[14] The indexes of similarity for five items in the 'neutral' section of the scale continuum were computed for each of the four scales. This made possible the construction of 20 diagrams, the general appearance of each one constituting the criterion of irrelevance for a given item with a 'neutral' scale value. These 20 diagrams should be, if the criterion is to be satisfied, in the form of an inverted U. The indexes should be high for items with scale values toward the center of the continuum and should decrease for items with scale values toward both extremes.

Not all 20 diagrams have been published, but the ones shown here are typical of the others. Obviously, the picture is not as it should be according to the criterion of irrelevance. All of these items would be rejected and would not appear in the published scales.

Following Thurstone and Chave, we can only conclude that in the case of diagrams such as Figs. 1, 2, and 3, that individuals who endorse these items "are just as likely to indorse statements at either extreme of the scale as the statements that are scaled near to" these items (19, p. 50). "This indicates that there is something fundamentally wrong with" the statements (19, p. 50). If the individual endorses these items we can say nothing about his attitude toward the issue in question because he is just as likely to endorse items at both extremes of the scale continuum, so far as can be determined by the statements under consideration. The endorsement of these statements "therefore does not help us in allocating the subject to a point on the scale" (19, p. 51).

Thurstone and Chave interpret a diagram such as Fig. 4 as representing an unsatisfactory item also. For one of their items with a similar criterion of irrelevance, they state:

"People all over the scale indorse this statement although they may have quite different feelings or ideas in doing so. The mere indorsement of this statement does not help us in locating the voter on

[13] That the application of the criterion of irrelevance involves a tremendous amount of labor perhaps accounts for the infrequent mention of its use by those who have constructed scales by the method of equal appearing intervals. Most investigators who have used this method have apparently been satisfied with a determination of the *Q* and scale values of the items.

[14] Professor Thurstone, in response to a personal inquiry, later informed me that the criterion had not been applied in the construction of all of the scales put out under his direction and that he did not consider the criterion to be completely successful. "The factorial methods," he adds, "are much better for solving the same problem. The statements in an attitude scale should represent a one dimensional system and consequently the items should be so selected that their intercorrelations are of rank one. In factorial language the uniqueness of each item is what I formerly called the coefficient of irrelevance of an item in an attitude scale" (21). See also the discussion by Ferguson (3).

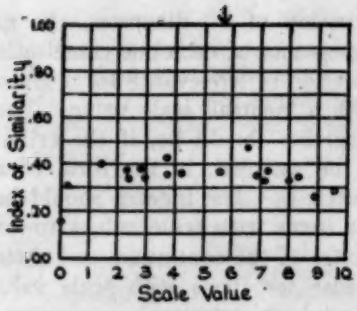


FIG. 1. Censorship, Item 13 (5.7). Whether censorship is good or not depends entirely upon the censor.

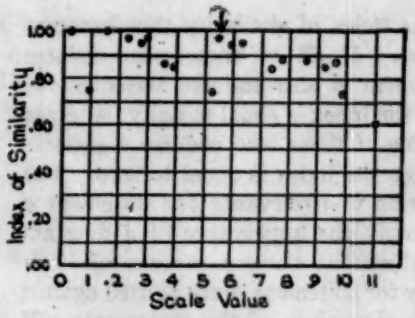


FIG. 2. Capital Punishment, Item 19 (5.7). It is unfortunate that we have no efficient substitute for capital punishment.

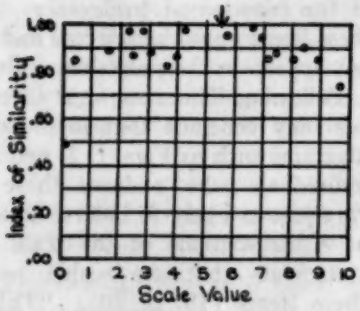


FIG. 3. Censorship, Item 5 (5.5). There is much to be said on both sides of the censorship question.

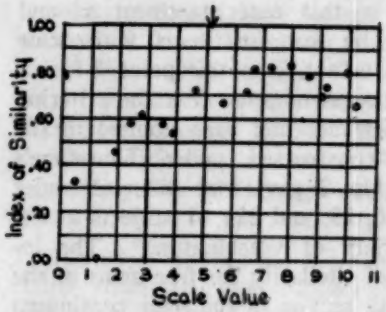


FIG. 4. Evolution, Item 10 (5.1). The evolutionary theory sounds logical but I don't know if it's true.

the scale. The statement is therefore judged to be irrelevant to the scale, which is represented by the whole list of statements" (19, p. 55).

Accepting the interpretations of the criterion of irrelevance advanced by Thurstone and Chave, we must conclude that in the case of the 'neutral' items examined the criterion is not satisfied.[15] An examination of the wording of some of these 'neutral' items will show some of the difficulties. Scale values of the items are given in the parentheses.

[15] Since not one of the 20 items which I analyzed satisfied the criterion, it is proper to raise the question as to the possibility of obtaining 'neutral' items which will prove to be satisfactory.

It would not be too difficult, I think, to picture the mental processes which might lead individuals with attitudes properly scaled at opposite extremes of the continuum to endorse these items. But since the possibilities are many, I leave them to the reader's imagination and analysis. The fact is that endorsement of any of these items is only slightly more revealing, if at all, as to the attitude held by the subject than would be endorsement of the statement: "It rains a great deal in Seattle."

VII. 'NEUTRALITY,' INDIFFERENCE, AND AMBIVALENCE

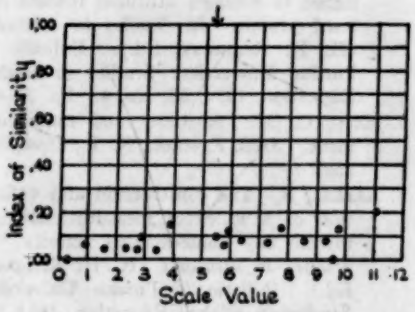
I mentioned earlier that, in addition to ambiguous and irrelevant items, indifferent and ambivalent items might

also tend to be sorted into the middle categories of the Thurstone continuum. The following examples, taken from two different Thurstone scales, indicate that this is the case.

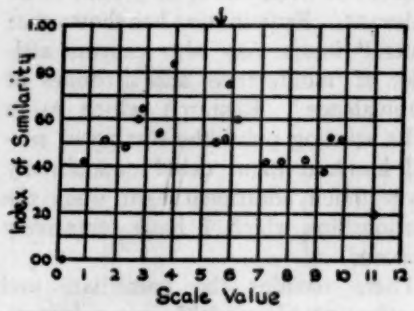
Despite the numerical similarity between the scale values of these pairs of items, do they not represent psychologically different states of mind? The (*a*) item in each pair, I believe, expresses an attitude of indifference or lack of concern about the issue, whereas the (*b*) item in each pair expresses a state of indecision or ambivalence.

We may observe also (from the criterion of irrelevance) that endorsement of the (*a*) items is not apt to be associated with a very great likelihood of endorsement of additional items, whereas endorsement of the (*b*) items, the ambivalent items, is associated with a higher probability of endorsement of other items. An individual who checks one of the indifferent items, in other words, is not as likely to agree with other items, either favorable or unfavorable, as the individual who endorses one of the ambivalent items.[16] This is precisely what we would expect if the (*a*) items are indicative of a state of

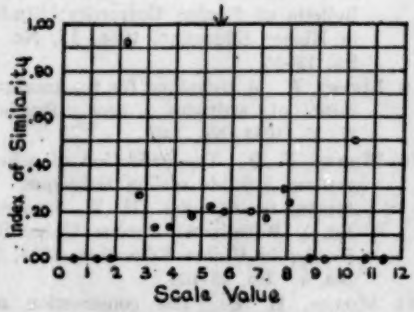
[16] On the scale measuring 'Attitude Toward Censorship,' Item 5 is also indicative of a state of indecision. The item, which carries a scale value of 5.5, is: "There is much to be said on both sides of the censorship problem." The criterion of irrelevance (Fig. 3) also shows that the likelihood of endorsement of additional items is high for those who agree with Item 5. Unfortunately, for purposes of comparison, this scale does not include an item which would signify indifference to the issue.



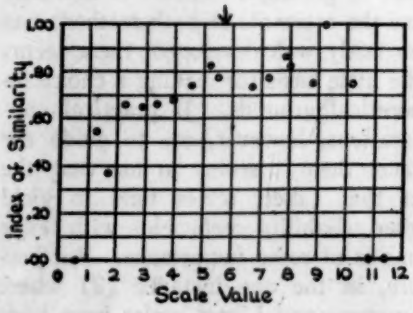
(*a*) FIG. 5. Capital Punishment, Item 1 (5.3). It doesn't make any difference to me whether we have capital punishment or not.



(*b*) FIG. 6. Capital Punishment, Item 6 (5.4). I don't believe in capital punishment but I'm not sure it isn't necessary.



(*a*) FIG. 7. Communism, Item 3 (5.5). Communism would not make much difference in the long run.



(*b*) FIG. 8. Communism, Item 16 (5.8). It is too early to judge communism by its results in Russia.

indifference and the (*b*) items of a state of ambivalence.

VIII. SUMMARY OF THE ARGUMENT

Although I have not made an exhaustive analysis of all of the 'neutral' items included in all of the scales constructed by the method of equal appearing intervals, I think that the evidence and examples cited indicate that some consideration must be given to the interpretation of these items. The fact that 'neutral' items tend to be non-differentiating seems to be established. We can only advance certain tentative hypotheses, at the present time, as to why they are non-differentiating.

In terms of the criterion of ambiguity, items in the 'neutral' zone tend to be relatively ambiguous. In terms of the criterion of irrelevance, they tend to be irrelevant. Examination has shown that 'neutral' items may also express attitudes of 'indifference' and attitudes of 'ambivalence.' Research which takes as its starting point the sorting of several hundred items could undoubtedly throw much additional light upon the relationships which I have tentatively proposed.

There remains the immediate and practical problem which many investigators must face: whether to use the method of equal appearing intervals or the method of summated ratings as a means of getting a measure of attitude. Since the rationale of both methods has been fairly well developed, there seems to be little basis for making a choice on theoretical grounds. If practical considerations, however, are to guide the choice, then it seems to me that the fact that Likert scales tend to yield higher reliability coefficients with fewer items is of some importance. Furthermore, in the one instance (2) where Thurstone and Likert scales have been independently constructed from the same set of original items, the correlation between scores on the two types of scales was found to be quite high. These two facts, in conjunction with the general agreement that Likert scales are much less laborious to construct, would seem to favor the method of summated ratings.

BIBLIOGRAPHY

1. BIRD, C. *Social psychology.* New York: Appleton-Century, 1940.
2. EDWARDS, A. L., & KENNEY, K. C. A comparison of the Thurstone and Likert techniques of attitude scale construction. *J. appl. Psychol.*, 1946, 30, 72–83.
3. FERGUSON, L. W. The requirements of an adequate attitude scale. *Psychol. Bull.*, 1939, 36, 665–673.
4. ——. A study of the Likert technique of attitude scale construction. *J. soc. Psychol.*, 1941, 13, 51–57.
5. GRICE, H. H. The construction and validation of a generalized scale designed to measure attitudes toward defined groups. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 37–46.
6. HALL, O. M. Attitudes and unemployment. *Arch. Psychol., N. Y.*, 1934, No. 165.
7. KELLAR, B. The construction and validation of a scale for measuring attitude toward any home-making activity. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 47–63.
8. KELLEY, I. B. The construction and evaluation of a scale to measure attitude toward any institution. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 18–36.
9. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.*, 1932, No. 140.
10. MILLER, F. D. The validation of a generalized attitude scaling technique. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 98–109.
11. MILLER, H. E. The construction and evaluation of a scale of attitudes toward occupations. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin

of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 68–76.

12. MURPHY, G., & LIKERT, R. *Public opinion and the individual.* New York: Harper, 1938.
13. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
14. PETERSON, R. C. *Scale for the measurement of attitude toward capital punishment: Form A.* Chicago: Univ. Chicago Press, 1931.
15. REMMERS, H. H. (Ed.). *Studies in attitudes.* Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4.
16. ROSANDER, A. C., & THURSTONE, L. L. *Scale for the measurement of attitude toward censorship: Form A.* Chicago: Univ. Chicago Press.
17. RUNDQUIST, E. A., & SLETTO, R. F. *Personality in the depression.* Minneapolis: Univ. Minnesota Press, 1936.
18. SILANCE, E. B., & REMMERS, H. H. An experimental generalized master scale: A scale to measure attitude toward any school subject. In *Studies in attitudes* (H. H. Remmers, Ed.). Bulletin of Purdue University: Studies in Higher Education, 1934, 35, No. 4. Pp. 84–87.
19. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Chicago: Univ. Chicago Press, 1929.
20. ——. *Scale for the measurement of attitude toward communism: Form B.* Chicago: Univ. Chicago Press, 1931.
21. ——. Personal communication. Dated February 19, 1945.
22. THURSTONE, T. G. *Scale for the measurement of attitude toward evolution: Form B.* Univ. Chicago Press, 1931.

SOME CONSEQUENCES FOR HISTORY AND PSYCHOLOGY OF LANGMUIR'S CONCEPT OF CONVERGENCE AND DIVERGENCE OF PHENOMENA

BY IVAN D. LONDON

Northwestern University

I. Introduction

In a previous paper it was pointed out that three more or less unrelated forms of indeterminacy are to be encountered in psychology: that arising from (1) standard difficulties characteristic of any act of measurement, or (2) inadequacies (inherent or otherwise) in the theoretical structure of psychological formulations, or (3) presence of divergent chains of events whose first links are essentially not ascertainable or duplicable (29, p. 168).

It was pointed out that, if not in practice, at least in principle, indeterminacy could be expected to give way to more determinate forms in psychological situations if those situations are characterized by indeterminate features identifiable as belonging to the first two types; and, further, that psychological situations characterized by divergence, as in type 3, present us with an impasse from which there is no escape as far as reduction to determinate forms is concerned (29, p. 168).

The latter conclusion stems from that aspect of physical quantum theory which makes predictivity of single subatomic events in principle impossible (29, pp. 162–164). Essentially on this account, I. Langmuir was led to propose a classification of sequences of phenomena which is singularly illuminating (22, pp. 3–4). He has proposed a classification such that sequences of phenomena are to be regarded as either convergent or divergent. They are convergent if the unpredictable initiating event sets off in conjunction with other unpredictable events a chain of events that in the aggregate over a period of time permits an averaging out into anticipated macroscopic behavior, thus converging, as it were, to a given result as if in response to the requirement of a definite law; that is, a definite law in the classical sense which renders the formulation: if A, then B, one that is always valid.[1]

They are divergent if the unpredictable event sets off a chain of events that, instead of converging to any anticipated macroscopic outcome, increasingly diverges with time into events quite unrelated to expectations. In the latter instance the classical formulation: if A, then B, fails to maintain operational validity and must be discarded. In Langmuir's own words:

". . . we must recognize two types of natural phenomena. First, those in which the behavior of the system can be determined from the average behavior of its component parts and second, those in which a single discontinuous event (which may depend upon a single quantum change) becomes magnified in its effect so that the behavior of the whole aggregate does depend upon something that started from a small beginning. The first class of phenomena I want to call *convergent phe-*

[1] This would be expected through Bohr's correspondence principle which implies that the classical laws of physics are simply the limiting forms assumed by the indeterministic formulations of quantum theory when the number of quanta or particles involved is very large. The molar laws of classical physics are then merely the simplified forms to which the statistical laws of quantum physics converge (13, p. 29; 29, p. 163).

nomena, because all the fluctuating details of the individual atoms average out giving a result that converges to a definite state. The second class we may call *divergent phenomena*, where from a small beginning increasingly large effects are produced. In general then we may say that classical physics applies satisfactorily to convergent phenomena and that they conform well to the older ideas of cause and effect. The divergent phenomena on the other hand can best be understood on the basis of quantum theory of modern physics (22, pp. 3–4).

Langmuir's concept of convergence and divergence of phenomena leads to interesting consequences for history as well as for psychology.[2] It is the purpose of this paper to trace some of them.

II. CAUSALITY

Before we proceed to the main body of discussion, a somewhat lengthy, but necessary, digression into the problem of causality must be made so that certain words and ideas in what follows may uniformly and unambiguously convey to the reader the specialized meanings that are intended. We are, of course, here venturing upon controversial ground where it is difficult to proceed entirely free from biassed emphasis of one sort or other. However, disagreements may be minimized if certain things are understood in advance and adhered to during the course of the subsequent discussion.

In the establishment of an operationally attested concept, there are always elements of convention to be considered; that is to say that, while the form of any concept may be suggested by the experiences to be covered by the concept, it is never completely determined by them (11, p. 440), since, it is to be emphasized, both the form of a concept and the way it is used are partially determined by the conventions imposed. The adoption of certain conventions, however, is not to be regarded as wholly arbitrary, for in the last analysis the conventions which are adopted are just those which wide experience shows are advantageous for the analysis of a situation (27, p. 526).

The concept of causality is no exception. Its form is determined by both experience and definition. In this paper we shall insist upon a form of the concept of causality which is operationally tenable. We shall, therefore, present causality in its empirical aspects as well as in those aspects of convention which seem to impute to it *a priori* status.

If we set up the relationship: if A, then B, as a starting point for a definition of causality, we are to regard the relationship as one that must be invariant throughout the flux of variable concomitancies that beset it. Invariance implies repetibility, which in turn implies the operational possibility of resetting the initial conditions covered by A (27, pp. 523–524).

By A or B we shall mean either an event or some aspect of an event which lends itself to the formulation: if A, then B. Context will in every case determine which is meant. By 'aspect' shall be meant *any* of the totality of perceptible properties of a given system or *any* of those numerous properties that indicate how any system which possesses them would behave when brought into certain problematic situations. The latter half of the definition of 'aspect' includes those many situations which phenomenally rarely repeat or are beyond the control of the observer, as the movements of heavenly bodies and their never-duplicating rearrangements in the celestial sphere.

[2] Some of these consequences have been anticipated on other grounds by the Neo-Kantians: Windelband and Rickert, as well as by others more recent, some of whom are noted in the references.

The presupposition in this case is that the aspects significant for invention of causal relationships are not entirely those which stand in easy revelation, but are also those which are intruded into phenomena by man's theorizations, as when he mathematizes or contrives a 'deep' explanation about them. And, finally, if the given aspect is sufficient to characterize the whole system completely, it will be said to define the 'state' of the system (9, pp. 14, 117; 28, pp. 269–270).

Two points must now be made. (1) For a system to obtain where such invariable formulation as, if A, then B, is possible, that system must be 'closed'; that is, cut off from varying influences or forces impressed from outside it. The concept of causality may be applied to an 'open' system only if the system can be closed by inclusion into systematic analysis of *those parts* of a larger system which produce the variable aforementioned 'impressing forces.' (2) It is *by definition* that the concept of causality is applicable to a closed system, because we call a system closed when its laws can be so formulated *independent of time* that the relation: if A, then B, invariably holds (27, pp. 522–523).

Furthermore, if the concept of causality does not apply according to a particular definition or formulation of a system, one simply redefines or reformulates what is meant by the 'state' of a system so that the concept may again be applicable. But, if such is the case, it is evident again that the concept of causality serves merely to define the term 'state.' From this standpoint causality is seen to be purely a conventional definition of what is acceptable of a state and the correlative theoretical system which attempts to provide a consistent tie-up of state A with a subsequent state B (27, p. 526; 28, p. 270).

Along these lines Lindsay and Margenau in their book on the foundations of physics state

> "By its definition, a closed or independent physical system is a causal one, because we call it closed when the laws governing its behavior do not involve time. But strictly speaking only closed systems are accessible to physical analysis. Thus, it would seem that physics can never inform us of a failure of the causality principle. . . . whenever a physical system does not appear to be closed, that is, when the differential equations describing it contain the time explicitly (if it does not behave consistently), we conclude that the variables determining the state in question are not completely known. We then look immediately for hidden properties whose variation may have produced the inconsistencies, and whose inclusion in the analysis would eliminate them; moreover, if we do not find any we invent them. This procedure is possible because in the consistency formulation of causality [by which nature is described in terms not containing explicit functions of time (27, pp. 522–523)] the term 'state' is [deliberately left] undefined" (27, p. 525).

So far, the emphasis in our discussion has been on the definitional implications of the concept of causality. The empirical and operational aspects must now be examined. Whenever we speak of the strictly invariable relationship: if A, then B, it must be realized that the relationship is one between idealized states, since it must be readily granted that actual states never correspond exactly to the conditions assumed for the projected relationship. Formulations relating states to one another have meaning only in so far as the former can be connected with verifiable results concerning the latter. Therefore, statements about a strict causality must be translatable into those about observable relations if causality is to have a utilizable meaning.

Following Reichenbach, this may be

accomplished as follows: states formulated in observational terms are always characterized by associated probabilities; so that we must really say, if A′ and its associated probability, then B′ and its associated consequent probability. In other words, we have a probability implication $A' \underset{p}{\supset} B'$ which states that, if A′ and its associated probability are given, then B′ will happen with a determinate probability p (11, p. 443).

From this empirically verifiable implication, we may surmise an ideal relation, which we call a law, by considering idealized states A and B and stating the logical implication $A \supset B$ between them. Since only with some probability can the ideal states A and B be inferred respectively from the observational states A′ and B′, a law expressing, if A, then B, is meaningful only in the sense of a transition to a limit. In other words, the possibility of going to the limit implies that the probability p can be made to approach the value 1 as closely as desired by increasingly intensive analysis of the characterizing variables involved in the description of states A′ and B′ (37, pp. 1–3).

With this understanding the concept of causality acquires a complete operational meaning. To say that elements of a situation are governed by a strict causality is then merely to say that they can be figured with a determinate probability and that this probability can be pushed as close to certainty as we wish by a sufficiently elaborate analysis of the phenomena under consideration. From one standpoint we have already seen that the concept of causality has almost *a priori* status. But, if the principle of causality is to be regarded, in addition, as a limit of probability implications, it becomes more than just a definition; it becomes also an empirical hypothesis. It is these two aspects which we must keep in mind.

The invariable relation: if A, then B, is not to be construed as implying any idea of physical compulsion or necessity (27, pp. 19, 192–193, 517, 520; 28, pp. 270–271; 44, pp. 225–226, 234), nor is any notion of temporal order involved (4, p. 118; 21, pp. 509–510; 25, p. 237; 26, p. 577; 28, pp. 274–275). Subsumed under this formulation are the three time orders: (1) where A precedes B, (2) where B precedes A, and (3) where A is simultaneous with B. H. M. Johnson, for reasons that will become apparent later, terms them respectively *mnemic*, *telic*, and *efficient* causation (21, p. 510).

Efficient causation implies a simultaneous ordering of A with B, and it is a harmless custom of speech to single out one as the physical cause of the other. Efficient causation is almost exclusively used in classical physics. For example, Hooke's law states that, within the elastic limits of a spring, extension is practically proportional to the force associated with it. Popularly, the force is said to cause the extension; but, as far as Hooke's law itself is concerned in its symbolic form, it is just as sensible to call the extension the cause and the force the effect as the other way around (28, p. 270).

As far as its symbolical relationships are concerned, classical physics is actually *a*causal in character if physical instigators or intermediaries from among concurrent symbolized events are insisted upon in one's concept of causality. In fact, it was only the advent of relativity that made causality in this sense at all possible, for space-time is constructed largely by means of causal laws which go on the assumption that, if there is a causal law connecting two events at different places in space-time, then they are connected by means of a chain of events in the intermediate

places (the well-known 'world-lines') (27, p. 520; 43, p. 427). Every event becomes, then, both recipient and instigator of action, since every event is connected with prior and future events in the space-time continuum by means of the aforementioned world-lines.

Mnemic causation implies that A precedes B, and is the type of causality which most people think of. Mnemic causation seeks to explain a given situation B by recourse to factors A in its past. Even physics has occasion to use such causation as in hereditary mechanics (27, pp. 55–56) or in dealing with hysteretic phenomena in general.[3]

In telic causation explanation of a given set of events is sought by appeal to a subsequent set. The minimal principles of classical physics make free use of telic causation (21, pp. 511–512; 27, pp. 116, 195). A light ray, for instance, takes the path it does, because it always travels that route which minimizes time of journey.[4] If a ray of light is at a particular spot on the way from its source X to its future point of impingement Y, then it is there because it is on that road from X to Y, which when completed in the future will turn out to have involved the least time.

If one is obsessed with the necessity of imputing a physical compulsion to causality, then in the above instance one would have to assert that the future *physically* determines the past, which conclusion most of us would find repugnant. But we have emphasized that the relationship: if A, then B, does not involve any notion of time order so that one could as well enunciate the relationship, if B, then A, without any danger of being misunderstood even though A temporally precedes B in the latter relation; *provided that* in the relationship: if A, then B, A is a necessary condition for B and B is sufficient for A.[5] Johnson drives this point home with this example:

". . . divorce D of a couple occurs after their marriage M; nevertheless the divorce is sufficient for the marriage: . . . , it is impossible that D should occur if M does not occur. I grant that it violates a language habit to say that divorce causes marriage, but the violation is not a violation of principle" (21, p. 511).

The preceding discussion of causality in its several aspects will, I trust, assist in keeping clear what it is that we are talking about in what follows and also assist in determining what conclusions may validly be drawn as to the combined effects of convergence and divergence of phenomena on history and psychology.

III. History

Since historical, as compared to natural, events are maximally divergent, historical events represent in an extreme form the divergence characteristic of much of psychology (8, p. 122; 22, pp.

[3] We merely allude to hysteresis here. We shall later discuss it more pertinently.

[4] This is known as Fermat's principle of least action which states that the actual path taken by light in traveling from one point to another irrespective of the medium is such that the time is less than for any other path connecting the two points; or, in the language of the calculus of variations,

$$\delta \int \frac{ds}{v} = 0,$$

where ds = element of path and v = velocity of light which varies with the media to be traversed (27, p. 135).

[5] H. M. Johnson states ". . . we cannot explain events of the present in terms of events of the future, because we do not know what the future will be. However, if both events are now past, we can, and often do, explain the earlier in terms of the later, by showing that the later is sufficient for the earlier and the earlier is necessary to the later. . . . , when we make use of telic principles of explanation, we do not have to assume that the system in its earlier state was 'conscious of' what its later state would be" (21, pp. 511–512; 44, pp. 211–242).

4–5). By a study of the effects of divergence for history, it will be possible to trace their consequences for psychology with a minimal interference from convergent factors. Afterwards, concurrent convergencies may be investigated in the psychological situation and their discerned effects superimposed upon those deriving from what is divergent.

The procedure to be followed is, in fact, reminiscent of the 'ideal' case of physics where at the first stage of theoretical investigation not actual gases are studied, but 'perfect' gases; not actual motion, but 'ideal' motion. For example, in order properly to understand actual motion in which resistance due to friction is encountered, motion in the unreal abstract situation without the complication of friction is studied and its laws mathematically traced out. Such ideal motion represents in the extreme what could be expected if the friction were progressively diminished, so that in the limit frictionless motion results. The study of actual motion in which friction is retained is then simply based on prior study of the ideal case, and, because it is an extension of the latter, the former becomes tractable to both computation and fundamental analysis. It is this 'ideal' procedure that will be followed.

Experimental science in its attempts to ascertain causal connections between phenomena always strives to duplicate the original conditions from which a certain state of affairs has been observed to emerge in order to discern whether any invariable relationship exists between the former and the latter of the form: if A, then B. It is obvious that history must differ from ordinary experimental science in that for the study of the causes of historical phenomena it is impossible to make use of the standard procedure of such science. Because it cannot duplicate circumstances that have passed and which in the very nature of things must defy exact repetition, history, as an open and divergent system, cannot make use of the formulation: if A, then B, in the same way that the exact sciences can. It is, therefore, unavoidable that causation in history cannot mean what it does elsewhere, for causality in the sense discussed in the previous section is here self-evidently not applicable. Except in so far as they may be tolerated as a handy convention of speech transposed from the habits of daily life, to speak of *laws* of history in the above sense, then, is an indefensible extrapolation of ideas into a domain of affairs in which the conditions that give those ideas point and meaning are absent.

A similar state of affairs exists in much of psychology, where once an event has taken place and the consequence noted, the event A cannot be either in principle or in practice duplicated in order to ascertain the invariable evocation of the consequent event B. In other words, the causal relation: if A, then B, cannot be established. If it is impossible to recreate the situation from which observed results have emerged, then statements implying causal relationships in the sense discussed in the previous section are here also self-evidently quite without meaning. Causality of this kind is operationally invalid if A cannot be in principle set up as frequently as may be desired. The invariance of the relationship: if A, then B, must be one that is attainable.

What meaning can one give to causality in situations where once an event has taken place it is neither in principle, nor in practice duplicable? Langmuir's concept of divergence of phenomena requires the admission that there are chains of phenomena which cannot be reduced to causal patterns. Yet the

human feeling persists that, because an event has preceded another, the first must somehow have influenced the happening of the second. How may these two opposing ideas be reconciled?

This may be done by taking the view that all detailed and exact historical causation, and much of psychological, is one derivable from hindsight only. This requires the frank admission that past events which have yielded a divergent progeny down into the present have explanatory significance only through reflection back from the present and that such retrospective 'causation' of present events is acceptable only because in the light of the present they feel right when so explained.

The following discussion may serve to clarify this proposition. It is a truism that any particular cross-section of human history presents an infinitude of events. Yet a historian, in attempting to explain why certain events now take place, does not record all discernible past events—a manifest impossibility—, nor does he pick out for consideration events at random. How does he proceed? He isolates a few of them as significant, and, as every event is the initiator of several trains of consequences, he tracks down only those bifurcations of historical events which to him seem significant. In other words, he must employ a theory of selection of events by which isolation of events, significant for him, becomes possible, and he must employ it successively through time down to the present (12, pp. 215–216).

But how does the historian arrive at his theory of selection of significant events and its 'law' of successive application? Certainly, he cannot use the procedure of experimental science in an attempt to evoke an invariable relationship between a state of affairs A and a subsequent state of affairs B. He does the following: he notes the end results of historical chains of divergent events —that is, the present. These indicate to him the direction of search for events which he must dig from out the matrix of the past and which he may then call relevant and significant. Chains of events that die out or, rather, sum up to nothing of importance from the *present* point of view would accordingly be ignored and, hence, said not to have 'caused' the present given state of affairs. In other words, they do not *feel* significant because nothing in the present suggests any pertinence to the present. Only *those* events that in the present feel significant forge for the historian the chains of events of which they are the last links.

Any such theory of selection is preeminently a subjective one[6] (12, pp. 215–216; 38). It and its law of successive application will constitute one-half the historian's system of explanation of world events; constructions whose elements are the selected events, the other half (57, pp. 278–279). According to this view, the present is the criterion of the past, and the historian a backward looking prophet.

Paradoxically, then, the halls of history will always resound to the echoes of the present, and the implication must be accepted that any given age will always reinterpret the history of the world in the light of what it regards as significant for itself; or, more baldly and in the last analysis, to suit itself and the tempering *Zeitgeist*. The inevitable divergent character of historical events forces this conclusion. There can be no 'true' or absolute explanations of historical events. The past must always be interpreted, as it were, in terms of the present.

If in our present-day world economic

[6] Every science or department of knowledge has its own theory of selection. Eddington with much reason even ascribes a 'subjective selectivism' to physics (13).

events are of great concern to us, then we will look for an economic interpretation of the past. In other words, because our personal experiences have made us feel that economic events are the significant events of our lives, we will, accordingly, try to isolate such events or related events from out of the matrix of the past to construct our history. We attempt thus an explanation of the course of events, which proves satisfactory only to the extent that the present allows such an explanation for us to feel right.

If this is a world whose fate is to be irrevocably fashioned by science, then a history, where the gradual emergence of science as the potent force in the world is ignored, will be one which will not satisfy us. The feeling that science is important is one which derives from the divergent end product of past events which is our 'now.' This indicates to us, for the purpose of setting up relationships with the present, which events in the complex of the past are to be singled out and isolated as significant. They will most certainly be events relating to science in some way. Generals, kings, and battles will be ignored, and men, unknown in their time and for ages afterwards, will be resurrected as the 'truly' significant men of history.

This is a procedure marked by the exercise of hindsight and is indelibly affected by the nature of events which got that way because they diverged into the present from the past in such a fashion that any explanation in terms of 'if A, then B' is in principle impossible. This is what history is; and, although such exercise of hindsight is indispensable, invaluable, and frequently satisfying, the point must be made: *this is not science.*[7]

It is the exercise of hindsight or what we have proposed to call retrospective causation, therefore, which makes it possible for Hartung to argue the thesis that the history of science and philosophy reveals their 'partisanship' and identity with the prevailing class interests of their times (**17**, pp. 121–122, 133; **23**, pp. 310–311; **30**, pp. 237–280). It is retrospective causation which makes it possible for Sloane to assert that the atomism of 19th century science is but the reflection of the atomistic individualism of the then dominant bourgeois society (**49**, p. 222); and this because Sloane is willing to maintain for *now* and, hence, for the past the following thesis:

". . . , scientific abstractions reflect the existing social relationships. Scientific preoccupation with separate, isolated units or elements of phenomena is a characteristic thought-way of bourgeois society and its whole ideological super-structure in which the emphasis has been on the social and political atom, the individual" (**49**, p. 222).

At any given stage of history, as regards almost anything, forces and counterforces for status quo, reaction, and change or revolution exist in complex intercalation. By ingeniously devising the requisite theory of selection one can find what one will to prove any thesis. Hartung and Sloane establish for *themselves* their contentions because their instrument of selection is based on a dialectical materialism or some related philosophy (**17**, p. 122), and, it must be admitted, their instrument probes well; but their success for *others* in so interpreting history is only proportional to the latters' own feeling concerning

[7] That is to say, unless one so enlarges the definition of science as to include everything contemplated in the 'scientific spirit' regardless of the limitations imposed by the tradition and methodology of the more advanced sciences; in which case the definition, losing stringency, runs the danger of being watered down to a mere expression of approbation applied to subjects so contemplated (**48**, pp. 234, 238; **58**, pp. 569–570).

dialectical materialism, and the counter thesis could as easily be established.

As another instance of the equivocation of historical explanation: if Germany practices a barbaric militarism, why, of course, she could not do otherwise, for does not such a barbaric tradition run through all her history? We, accordingly, find it possible for Guérard, in a review of one of Vansittart's books (52), to maintain in the following words that historically

> ". . . Germany as a *nation* is committed to a mad policy of ruthless domination. Not Hitler alone is guilty, but an ever-deepening trend, from Frederick II to Bismarck, to William II, to Goebbels and Goering. I deeply regret that Vansittart is right. It would be infinitely more comfortable if he were not. But the Germans have proved him right, and I have to accept the evidence" (15, p. 111).

And it would as easily have proved possible to discern, concurrently diverging from the same past, another 'ever-deepening' and widening trend, from Germany's great scholars, scientists, and artists to inheriting contemporaries who, but for the course of events—always half-way erratic except in retrospect—might have remained in the ascendant. If Germany had presented herself to us as a peace-loving nation, cultivating all the arts of civilization, we would most assuredly have found among us those for whom Germany's Goethe, Schiller, and Lessing prove she could not have developed otherwise.

Because of retrospective causation, by which the present reveals *a* significant past and is in turn explained by it, Burtt can contend that the great difference which obtains between the world-view of medieval and modern thought is explainable in terms of a change in the metaphysical foundations accepted by the pioneers of modern physical science; that is to say, a change traceable to a substitution of the dominant Aristotelianism of the medieval ages by a view of the universe as fundamentally geometrical in structure, which view is then asserted to be derived from the Pythagorean and Platonist tradition (10; 46; 50, pp. 62–67; 51, pp. 5, 50).

In essential support of the position assumed in this paper, Strong takes the opposing view that metaphysical tradition then and since has been largely irrelevant to science, which has always had its own 'operational autonomy,' and adopts conceptions and procedures because of their methodological efficacy rather than because they are corollaries of any metaphysics or cosmology. For this reason, according to Strong, we should be wary of assuming too readily that the conclusions of physical science either substantiate or are dependent upon the metaphysical doctrines with which they may seem to have a clear affinity (51, pp. 8, 115–123, 214, 217).

The divergence characteristic of historical events makes either alternative view plausible. It is a matter of taste and inclination which view is adopted. Philosophers undoubtedly will cotton more quickly to Burtt's body of ideas; physicists toward Strong's. For our purpose, however, it is important to note that both men derive their explanations by the use of two opposing theories of selection of the significant factors of history. They both make use of retrospective causation; but Strong cautions against Burtt's indiscriminate use of it—an apt admonition, particularly in view of the fact that the modern interpretation of Plato, which Burtt evidently has in mind, is quite a different thing, as Strong well demonstrates (51), from the Pythagorean-infused mysticism that passed for Platonism in the medieval ages.

Bell shares the views of Strong so far as the intellectual autonomy of

most of mathematics is concerned. The ascription of the particular course taken by the development of mathematics to the motivation of economics and class interest finds Bell likewise in refutation stating,

"Several items . . . suggest that much of the impulse behind mathematics has been economic. In the third and fourth decades of the twentieth century, for obvious political reasons, attempts were made to show that all vital mathematics, particularly in applications, is of economic origin.[8]

"To overemphasize the immediately practical in the development of mathematics at the expense of sheer intellectual curiosity is to miss at least half the fact. As any moderately competent mathematician whose education has not stopped short with the calculus and its commoner applications may verify for himself, . . . , it simply is not true that the economic motive has been more frequent than the purely intellectual in the creation of mathematics. This holds for practical mathematics as applied in commerce, including all insurance, science, and the technologies, as well as for those divisions of mathematics which at present are economically valueless. Instances might be multiplied indefinitely; . . ." (2, p. 20).

[8] In Russia it is still the practice of many to appeal to Marx, Engels, and Lenin as authorities in matters obviously outside the latters' competence. A recent example is Voronkov, who maintains that the history of any science, and that of mechanics in particular, proves its development derives largely from the technological and economic circumstances of the times and as corroboratory witness quotes Engels to the effect that science is by far more bound to the state and level of development of the means of production than the other way around (53, p. 10). On the other hand, however, Zhebrak condemns the careless application of dialectical materialism to scientific matters and, as an instance, strongly disapproves the position of Lysenko concerning the latter's supposed refutation of Mendel's laws of heredity on the basis of it (59, p. 358; also L. Kartman, Soviet genetics and the 'autonomy of science,' *Sci. Mon.*, 1945, 61, 67–70).

Before relinquishing discussion of historical causality, it might be well parenthetically to note that the divergence of historical events lends confidence to the feeling that the future cannot quite be the slave of the past; that in a certain sense the present always represents a semi-break with the past (8, pp. 29–30, 32), whose converging aspects dull the sharpness of the schism between them; that perhaps man, by taking advantage of the appearance of favoring divergencies, can help fashion his own future and prejudice the fates in his behalf. By so doing man can tip the scale of events his way (16, p. 132).

IV. PSYCHOLOGY

So far, we have dealt with the study of the events of mankind; we turn now to the study of the events of *one* man. Is he amenable to scientific reduction, or will he present the same difficulties to the psychologist as history to the historian? If the latter, then how much of him, nevertheless, can be made subject to the discipline of a scientific psychology?

The preceding discussion allows us now clearly to perceive that a situation, similar to history, obtains for much of human phenomena which psychology would attempt to incorporate into an exact science. And this because the presence of divergent phenomena everywhere—in man as well as in nature—makes reduction in explicit detail of psychological phenomena to the invariable relationship: if A, then B, an effort that will not meet with constant or universal success. 'Causal' relationships that are noted will bear the mark of retrospective causation, and, analogously to the historical situation, items of past behavior will prove significant only in the light of the reacting individual's present behavior as gauged by the observing psychologist.

For instance, if sexual disturbances

are suspected of contributing to form a given pattern of behavior in an adult by a psychologist for whom sexual factors seem significant, then it becomes incumbent upon him to look for related behavior in the past of that individual, which when discerned will be called genetic or causative. Accordingly, a childhood history of sexual aberration in one individual will be considered as significant if that individual subsequently turns out abnormal; while a similar childhood history in another individual who, contrariwise, presents a picture of normal adulthood is considered as not 'truly' significant, and appeal made to 'other mitigating factors' and 'sublimations' to explain away the difference in outcome.[9] But this is the practice of hindsight, and it must be recognized that the presence of divergent phenomena in human beings prevents us from predicting the separate future behavior of two individuals presenting similar aspects of behavior in childhood.

Only probabilities may be hazarded, and these again are arrived at through present insight based on our own past experiences, divergently modified and modifying within and without us (9, pp. 160–161). We feel that such and such a situation is likely, we feel that such and such a therapeutic procedure will prove effective, we size up a situation by putting ourselves in it or an individual by putting ourselves in his shoes and noting the projected effect on ourselves. Because of this, many explanations, as in the dynamic psychologies, relating past behaviorial phenomena with present manifestations satisfy, because they feel right; and therapeutic procedures based on such explanations will have the frequent success that they do have, because we cannot keep ourselves out of them (5, p. 220; 18, p. 70).

It is for this reason that divergence of phenomena prevents much of psychology and psychiatry from being exact sciences in the sense that it is always possible to set up the invariable relationship: if A, then B; and gives them more the aspect of an art than a science[10] (6, p. 384; 58, p. 570). This is inherent in the situation, and need not on that account be deplored. It becomes the job of the psychologist as the scientist to find out how much of psychology may be reduced to the formulation of the invariable relationship: if A, then B, and what aspects of divergent behavior still are susceptible to such treatment. Much of human behavior is convergent and, hence, amenable to scientific reduction, but that which is divergent in character must remain indeterminate and be delivered over for appraisal to the psychologist's 'insight.'[11] In the latter contingency

[9] It may be objected that the very difference in individual outcomes contradicts the assumption of identical childhoods, but this objection is invalid, stemming, as it does, from the implicit belief in the universal convergence of phenomena and the use of non-operational definitions of causality. Retrospective causation, from the point of view of strict causality in the sense previously discussed, is a designation that is operationally undefinable, being merely a harmless, but convenient, verbal concession to the feeling that all things must somehow or other be related and that only ignorance prevents discovery of the relations.

[10] In this regard Rosenzweig recalls Dr. A. Angyal's witticism that "psychotherapy is the art of applying a science which does not yet exist" (42, p. 521).

[11] In harmony with this reasoning, Skaggs (48, pp. 234, 238) in concordance with Allport (1, p. 22) suggests the acceptance of Windelband's two-fold division of disciplines into the *nomothetic* and the *idiographic:* the former seeking only 'general laws' and employing only 'those procedures admitted by the exact sciences'; the latter endeavoring to 'understand some particular event in nature or society.' The argument as it has so far proceeded in this paper makes this division for psychology an agreeable one. It must,

'causal' connections will be hypothecated on the basis of subsequent events which we feel will precipitate from out the fluid past those factors which are significant for incorporation into an 'operationally' (!) retrospective theory.

In a sense very similar to that of the historian, the psychologist utilizes for discerning causal relationships over a long period of diverging behavior a subjective theory of selection of significant events, interpreting the past through the eyes of the present with the subsequent application of the former to explain the latter (28, p. 283). It must be stressed, however, that the psychologist is more fortunate than the historian in the choice of his theory of selection and that of its successive application, in that much of his domain of interest is reducible or is in principle amenable to reduction to the invariable relationship: if A, then B. We may conclude, then, that the more divergent in character the behavior, the more psychological theory becomes for the psychologist merely the facilitating rule or device for the successive application through time of the theory of selection which he has chosen.

Before going on to other related questions, it is again well to observe that, as in the case of historical events, the divergence of psychological events must lend confidence to the feeling that the individual is not quite subservient to the years that went before, nor is he the invariable expression of his childhood down to the last detail; that in a certain sense his future always represents a semi-break with his past, 'lawful' continuity being maintained by the converging aspects of those events for which he was both machine and repository; that a man, by taking advantage of the appearance of favoring divergencies, within and without, can help fashion his own future and semidirect his fate. For so doing, every moment presents its opportunity and renders him his own circumscribed free-agent.

According to this view, then, free-will may be looked upon as a function of divergence, operating within the limits imposed by the internally and externally convergent. Free-will is a term assignable to the feeling of latitude of choice that attends acts divergently characterized. It may be protested that this is a pseudo-freedom, an illusion. But, in view of the fact that the very divergence of acts seemingly covered by free-will makes any test of free-will impossible, this objection is pointless, for divergence forbids the possibility of resetting those conditions which obtained prior to the exercise of 'free-choice' (44, pp. 229–230; 47, pp. 21–22), in order to demonstrate either free-will or its illusion. It becomes a matter of personal predilection, whether one chooses to become excited one way or the other over the question. Free-will is merely a token of the divergence of events and as a convenient neutral term admissible to language (19, pp. 37–38).

V. HOLISM *vs.* ATOMISM

Langmuir's concept of convergence and divergence of phenomena allows us to take a fresh view of the current dispute between the holistic and 'atomistic' points of view in psychology. It will be clear from the succeeding discussion and from what has already preceded that the frequently bitter controversy in psychology between the holists and the atomists stems essen-

however, be pointed out for accuracy's sake that Windelband intended psychology to be counted as *durchaus* nomothetic, and only the 'historical' disciplines as idiographic (56, p. 26). Allport, while approving in general, finds the division of psychology into nomothetic and idiographic disciplines too sharp and prescribes their inclusion as mutually-assisting disciplines 'within the scope of a *broadened* psychology' (1, pp. 22–23).

tially from the use of different definitions of what constitutes a science and from a lack of appreciation by the former of the convergent nature of much of psychological phenomena and by the latter of the mushrooming effects of accompanying divergencies which can get so to dominate a psychological situation as to reduce whatever is convergent in it to a status of near irrelevancy.

Analysis of wholes into parts, as implied in atomism, for the purpose of scientific investigation—that is to say, for the purpose of discerning uniformities in wholes such as to allow the invariable kind of formulation: if A, then B, is predicated on the assumption that the phenomena with which it is dealing are convergent in nature; otherwise, A cannot ever be duplicated to test the invariancy of the relationship set up. Furthermore, wholes in the perceptual raw are not fit material for reduction to invariant law without some prior analytical processing.

It is for these reasons that simplification and isolation of phenomena are a *sine qua non* for the determination of causal relationships. Science has never been able to proceed otherwise, for disturbing elements or irrelevant divergent factors, when permitted free rein in experimental situations in order to preserve 'the total situation in all its concrete individuality' (24), can serve no other purpose than to delay or prevent the formulation of laws.

It has been remarked in another connection how fortunate it was for the world that Kepler did not have at his disposal more complete and refined data of planetary motion; for, if he had, he would never have been able in all the increased intricate complexity of astronomical data to discern the simple relationships which he was able compactly to express in his famous three laws and which helped give the impetus to further discovery by Newton (3, p. 568). If this can be said of Kepler who dealt with phenomena non-divergent in character and supremely susceptible of formulations of the kind: if A, then B, what can be said in defense of the holists [12] who deal with admixtures of phenomena, both convergent and divergent, without attempting to isolate that which is amenable to lawful construction from that which is not? Continually to assert that the atomists believe that 'the whole is equal to the sum of its parts'—the straw man in psychology that some holists are always knocking down—does not obviate the need of eliminating divergent factors wherever there exist psychological phenomena that promise to lend themselves to strictly scientific formulations.

However, in divergent situations, as in trying to understand a complicated subject as the human being in flux and strain of action, the holistic point of view becomes the valuable one; but this understanding is one not based on scientific laws of the kind: if A, then B, but one which might be called empathetic, and its individual success is directly proportional to the degree to which it feels adequate to any given situation. This has already been pointed out at length in the preceding discussions (5, p. 220).

Analysis into parts does not necessarily imply for resynthesis into the whole simple additiveness (14, p. 583); for any successful analysis must provide, not only the rules for systematic fragmentation of experience, but also those for its reknitting. On this point Eddington states,

> "For a scientific outlook I think the most fundamental of all forms of thought is the *concept of analysis*. This means the conception of a whole as divisible into

[12] This includes the Gestaltist, the organicist, and all variants.

parts, such that the co-existence of the parts constitutes the existence of the whole. . . . The point which I must emphasize is that I am referring to the conception of *a set of parts*, not to the individual conception of *a part*. In the concept of analysis a part is always a member of a complete set of parts, and its significance is bound up with the system of analysis in which it occurs" (13, p. 118).

If, therefore, the given rules for acceptable analysis are not successful, new ones must be sought. Too often, however, when this proves difficult, recourse is had to holism. But this is a confession of failure, so that in this instance holism may be properly defined as a function of ignorance. We still, for example, read statements by writers in defense of the holistic point of view that combinations of atoms in molecular form have different qualities or properties than do the constituent atoms (49, pp. 215–216). Of course, but who ever said that they did not? That is not the issue. The issue is this—are these properties which combinations of atoms exhibit derivable from a consideration of the properties of the individual atoms which compose them?

The holists say no, and as an instance assert, as if to prove their point, that hydrocarbons have qualities which are essentially different from those of the individual hydrogen and carbon atoms which make them up. But the point is again this: are the properties of the molecular whole derivable from a consideration of the properties of its parts —in this case, the individual atoms? The answer is yes (7, 34, 54), since the properties of hydrocarbons such as their heats of fusion, vaporization, hydrogenation, and combustion, their melting and boiling points, viscosity, density, color, refractive index, chemical affinities, absorption spectra, etc. are in point of fact to an increasing degree so deducible despite frequent incredible statements to the contrary (7, Chaps. I–V; 49, p. 216; 54, pp. 79–81, 158).

It is the increasingly successful reductionism of chemistry that is alluded to when Wheland, in speaking of the theory of resonance, 'the most important addition to chemical theory since the introduction of the concept of the shared-electron bond,' states

"Its basis lies in the mathematical depths of quantum mechanics. For that reason, it can be presented precisely and completely only in highly mathematical language. . . . Experience has shown, . . . , that often more difficulties are created than are avoided if the attempt is made to ignore entirely the underlying physical basis of the theory. . . . This new quantum mechanical approach not only provides a sound theoretical basis for . . . earlier and less precise ideas [in chemistry] but also supplements them in several important regards by introducing in them certain refinements or extensions which could have been reached in other ways only with great difficulty. . . . We are now in [the] position to consider the question of the effect that resonance has upon the physical and chemical properties of the molecules in which it occurs" (54, pp. iii, 5, 28).[13]

[13] To assert with the holists that each 'level of organization' of phenomena *demands* its own peculiar set of laws as distinct from any other is to be, as regards the future development of science, merely plausible; as regards its past development, simply uninformed. The history of science right up to the recent instance cited above demonstrates not only that its advance has been a recurring lesson in reductionism (generalization from one level to another more fundamental), but also that the scientist's faith in his ability to so generalize is not a vain one so long as it is nomothetic material that he is handling.

This does not mean that there is no place for limited laws particularly adapted to a restricted area within a given level of organization. After all, to solve an ordinary problem in engineering one does not have recourse to relativity and quantum mechanics, but to the more convenient classical variety which the former through increased scope and power of generalization displaced.

Strictly speaking, on a more fundamental level one does not ask whether the properties of hydrocarbons are reducible to the properties of the constituent atoms, but, rather, the question, like all such questions, should be operationally put—may the properties of hydrocarbons be calculated by means of the same basic methods by which were deduced the properties of the constituent atoms? It is only if this is kept in mind that it is possible for one to appreciate to what lengths the reductionism of chemistry has really proceeded, when Wheland observes that

". . . [the theory of] resonance is a man-made concept in a more fundamental sense than most other physical theories. It does not correspond to any intrinsic property of the molecule itself, but instead it is only a mathematical device, deliberately invented by the physicist or chemist for his own convenience" (54, p. 28).

In other words, reverting to our less sophisticated manner of speaking, the proper rules for resynthesis of the parts into the whole in the case of the hydrocarbons are available. It is apparent that, for convergent phenomena, prior to such successful formulation of rules for resynthesis of parts into wholes, the willingness to adopt a holistic explanation, which from the point of view adopted in this paper in reality is no explanation, but mere phenomenal description, is a direct function of one's ignorance and failure to find the right rules to do the job demanded of acceptable analysis (5, pp. 214–215, 217–218; 28, p. 279; 31; 32, pp. 209–210, 214; 33).

VI. HYSTERESIS

The indeterminacy encountered in so much of psychology was in a previous paper (29) found fundamentally traceable to three sources: (1) difficulties in measurement, (2) inadequacies in theoretical structure, (3) divergencies (29, p. 168). H. M. Johnson has, however, directed my attention to an alternate way of viewing the indeterminacy that inheres in much of psychology. He argues that indeterminacy depends on the fact that living organisms constitute hysteretic systems, in which every response sets up mnemic traces in their structure, so that the next stimulus impinges on a system that may differ quite markedly from the original. Consequently, the act of producing a response may so change the system that the second stimulus, though externally equivalent to the first, produces a different response. Johnson points out that this fact was known to Fechner who was on that account driven to define the concept of threshold statistically and ventures, further, the opinion that this concept—the most fundamental in 'classical psychology'—is one that is *necessarily* statistical because of the hysteretic character of the behavior of living organisms.

As an example, Johnson adduces the selective reaction-time situation where multimodal frequency-distributions are always obtained. After as many as 5,000 reactions have been taken, even if the times are recorded simultaneously on two chronoscopes, one measuring in units of 0.01 second and the other in units of 0.001 second, multimodal frequency-distribution persists whether the instrument is set back toward zero or not. This is for Johnson an example of the indeterminacies that attend every psychological measurement. He emphasizes that they are due not to the fault of the experimenter or to the crudities in his method of measurement, but exist because the human being behaves as he does—that is, because he is a hysteretic organism.

Hysteresis may be defined very generally as being the dependence of the properties and behavior of a system in theoretical formulations on not only the

instantaneous environmental conditions, but also the entire previous history of that system (35, pp. 208, 214; 36, p. 368; 47, pp. 17–18). A piece of iron, for instance, which has a complicated magnetic history may respond in a great variety of ways to alterations of the magnetic field to which it is exposed, even if its magnetic field and induction (the parameters which are usually considered to determine the magnetic state of a body) have been assigned values. Under such circumstances the piece of iron will be said to exhibit magnetic hysteresis and behavioral indeterminacy (9, pp. 17, 61–62).

This would seem to bolster Johnson's thesis that hysteresis is the significant explanation for psychological indeterminacy, were it not for the fact that hysteretic phenomena have frequently shown themselves reducible to determinate forms, either theoretically or actually.

Bridgman makes the supposition, for instance, that the state of an iron body exhibiting hysteresis cannot be described by the large-scale measuring operations which assuredly give the ordinary characterizing parameters, but which fail to disclose its fine structure. For this latter purpose it would be necessary to determine the characterizing parameters of all the individual elementary magnetic domains which make up the body. In principle this can be done, although the actual experimental justification is beyond the practical stage at the present[14] (9, pp. 63–64, 136). This, however, leads hysteresis back to the first of the three enumerated types of indeterminacy discerned in psychology.

Bridgman also points out that for bodies showing hysteresis it is frequently possible for the solution of certain problems to apply the energy concept provided the bodies are reduced to some standard state which does not show hysteresis as by dissolving a magnet in acid and, further, that certain classes of alloys become entirely manageable and lose their hysteretic character simply by the introduction into theory of a single new parameter measuring molecular 'disorder' (9, pp. 63–64). This, however, leads hysteresis back to the second of the three enumerated types of indeterminacy, and underscores the fact that this type of indeterminacy lies not so much in the phenomena involved, as in the viewer of the phenomena: *it is relative to the adequacy of his explanatory theory* (5, p. 211; 14, p. 585; 20, pp. 489, 491, 493).

Since we have already shown one instance of hysteresis reducible to type 1 and another to type 2, the conclusion must be drawn that only when hysteresis is in principle irreducible to non-hysteretic formulations is it in principle indeterminate. It is then equivalent to the indeterminacy of type 3 which refers to divergence of phenomena.

Hysteresis in psychology assumes that 'cross-sectional' laws (efficient causation), directly connecting simultaneous psychological variables, in many given situations, 'open' or 'closed,' are impossible and implies that knowledge of the generating past of such psychological situations must be juxtaposed to knowledge of their present structures in order to allow for a tolerable extrapolation into their problematic futures (29, p. 167).

Why so? Let us recollect that the formulation, $A' \underset{p}{\supset} B'$, expresses an empirical relationship to the effect that, if A′ and its associated probability are given, then B′ will happen with a determinate probability p. The ideal formulation, $A \supset B$, which we call law, implies that p can be made to approach

[14] Johnson, himself, alludes to this possibility (21, p. 512).

1 as closely as desired by increasingly intensive analysis of the characterizing variables involved in the description of the states A′ and B′, so that $A \supset B$ is the limit of probability implications of the form, $A' \underset{p}{\supset} B'$ (37, pp. 1–3).

But it is important to note that, if causality is looked on as an empirical hypothesis, then there is no *logical* necessity that compels p to approach certainty by an increasingly fine analysis of states and thereby to bring about a transition to the limit which is expressed by the ideal relation, $A \supset B$. Furthermore, simultaneous determinations of variables in practice frequently become increasingly difficult as the number of the latter is increased and the analysis of them made finer, for the simultaneous determination of any number out of a given set of variables disturbs the remainder, preventing the attainment of any cross-sectional aspect in its required entirety[15] (29, p. 164; 37, pp. 3–4).

Recourse, then, must be had to sequences of uniformities discerned in the past of similar psychological situations, and hysteretic formulations (mnemic causation) hypothecated on the uniformites noted therein. Hysteresis in such cases affirms the impossibility of exact predictive laws solely based on *necessarily* incomplete knowledge of present psychological structure. In addition, divergence contrives to decrease further the degree of predictability which hysteretic formulations can attain. And this, as we now can see, is the real significance of hysteresis and divergence for psychology.

[15] The famous example of the limit to predictability despite all attempts at increasingly fine simultaneous determinations of the variables involved is the Heisenberg inequality, $\Delta x \cdot \Delta p \geq h/4\pi$, of quantum physics (29, pp. 162–163).

VII. CONCLUSION

Langmuir's classification of sequences of phenomena into those which are convergent in character and those which are divergent leads to consequences, some of which are summarized by way of recapitulation as follows:

(1) There are no such things as immutable laws of history, nor may they be called 'true' or absolute; every age having to interpret the past in the light of its present.

(2) Explanation in history is by hindsight only and proceeds by the application of a subjective theory of selection of events.

(3) Similarly, much of psychology is not amenable to scientific reduction, in the sense that invariable relationships of the kind: if A, then B, are attainable, applicable, or even desirable; retrospective causation playing here the leading role.

(4) The problem of free-will from the point of view of a strict operationism is a pseudo-problem, free-will being a function of divergence and, as such, admissible to language as a convenient neutral term.

(5) Holism is a valuable point of view when divergent factors are involved in situations to be evaluated.

(6) Atomism (analysis) is the appropriate point of view for the study of strictly convergent phenomena and of the convergent aspects of phenomena principally or somewhat divergent in character.

(7) Hysteresis is a concept that covers a range of phenomena which in principle may or may not prove determinate according as they are or are not covered by any one of the three proposed classes of indeterminacy; hysteresis remains, however, a valuable concept which even in its unanalyzed form illumines the indeterminacy characteristic of so much of psychology.

BIBLIOGRAPHY

1. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937. Pp. 588.
2. BELL, E. T. *The development of mathematics.* New York: McGraw-Hill, 1940. Pp. 583.
3. ——. Newton after three centuries. *Amer. Math. Mon.*, 1942, **49**, 553–575.
4. BERENDA, C. W. On Birkhoff's and Einstein's relativity theory. *Philos. Sci.*, 1945, **12**, 116–119.
5. BERGMAN, G. Holism, historicism, and emergence. *Philos. Sci.*, 1944, **11**, 209–221.
6. BILLS, A. G. Changing views of psychology as science. PSYCHOL. REV., 1938, **45**, 377–394.
7. BRANCH, G. E. K., & CALVIN, M. *The theory of organic chemistry.* New York: Prentice-Hall, 1941. Pp. 523.
8. BRIDGMAN, P. W. *The nature of physical theory.* Princeton: Princeton Univ. Press, 1936. Pp. 138.
9. ——. *The nature of thermodynamics.* Cambridge: Harvard Univ. Press, 1941. Pp. 229.
10. BURTT, E. A. *The metaphysical foundations of modern physical science.* New York: Harcourt, Brace, 1932. Pp. 343.
11. CARNAP, R. Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis*, 1931, **2**, 432–465.
12. CASSIRER, E. Neo-Kantianism. In *Encyclop. Brit.*, Vol. 16, 14th ed., 215–216.
13. EDDINGTON, A. S. *The philosophy of physical science.* New York: Macmillan, 1939. Pp. 230.
14. GERARD, R. W., & EMERSON, A. E. Extrapolation from the biological to the social. *Science*, 1945, **101**, 582–585.
15. GUÉRARD, A. Apologies to Vansittart. *Nation*, 1945, **161**, 111–112.
16. HALDANE, J. B. S. A new theory of the past. *Amer. Scient.*, 1945, **33**, 129–145.
17. HARTUNG, F. E. The social function of positivism. *Philos. Sci.*, 1945, **12**, 120–133.
18. HERRICK, C. J. The natural history of experience. *Philos. Sci.*, 1945, **12**, 57–71.
19. HEMPEL, C. G. Analyse logique de la psychologie. *Rev. Synthèse*, 1935, **10**, 27–42.
20. HENLE, P. The status of emergence. *J. Philos.*, 1942, **39**, 486–493.
21. JOHNSON, H. M. Rival principles of causal explanation in psychology. PSYCHOL. REV., 1939, **46**, 493–516.
22. LANGMUIR, I. Science, common sense, and decency. *Science*, 1943, **97**, 1–7.
23. LENIN, N. *Materialism and empirio-criticism.* Vol. 13 of *Collected Works of V. I. Lenin.* New York: International Publishers, 1927. Pp. 342.
24. LEWIN, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936. Pp. 231.
25. LEWIS, G. N. Light waves and light corpuscles. *Nature*, 1926, **117**, 236–238.
26. ——. The symmetry of time in physics. *Science*, 1930, **71**, 569–577.
27. LINDSAY, R. B., & MARGENAU, H. *Foundations of physics.* New York: Wiley, 1936. Pp. 537.
28. LONDON, I. D. Psychologists' misuse of the auxiliary concepts of physics and mathematics. PSYCHOL. REV., 1944, **51**, 266–291.
29. ——. Psychology and Heisenberg's principle of indeterminacy. PSYCHOL. REV., 1945, **52**, 162–168.
30. MANNHEIM, K. *Ideology and Utopia.* New York: Harcourt, Brace, 1936. Pp. 318.
31. NEEDHAM, J. A note on Dr. Novikoff's article. *Science*, 1945, **101**, 582.
32. NOVIKOFF, A. B. The concept of integrative levels and biology. *Science*, 1945, **101**, 209–215.
33. ——. Continuity and discontinuity in evolution. *Science*, 1945, **102**, 405–406.
34. PAULING, L. *The nature of the chemical bond.* Ithaca: Cornell Univ. Press, 1942. Pp. 450.
35. RASHEVSKY, N. Learning as a property of physical systems. *J. gen. Psychol.*, 1931, **5**, 207–229.
36. ——. Possible brain mechanisms and their physical models. *J. gen. Psychol.*, 1931, **5**, 368–406.
37. REICHENBACH, H. *Philosophic foundations of quantum mechanics.* Berkeley: Univ. of California Press, 1944. Pp. 182.
38. RICKERT, H. *Der Gegenstand der Erkenntnis*, 1904.
39. ——. *Die Grenzen der naturwissenschaftlichen Begriffsbildung*, 1913.
40. ——. Geschichtsphilosophie. In *Die Philosophie im Beginn des zwanzigsten Jahrhunderts* (W. Windelband, Ed.), 1907, pp. 321–422.
41. ——. *Kulturwissenschaft und Naturwissenschaft*, 1899.
42. ROSENZWEIG, S. A dynamic interpretation of psychotherapy oriented towards research. *Psychiatry*, 1938, **1**, 521–526.

43. RUSSELL, B. *An inquiry into meaning and truth.* New York: Norton, 1940. Pp. 445.
44. ——. *Our knowledge of the external world.* Chicago: Open Court, 1914. Pp. 245.
45. ——. *The scientific outlook.* New York: Norton, 1931. Pp. 277.
46. SCHRÖDINGER, E. Ist die Naturwissenschaft milieubedingt? In *Zwei Vorträge zur Kritik der naturwissenschaftlichen Erkenntnis.* Leipzig: Barth, 1932. Pp. 62.
47. SEMON, R. *Mnemic psychology.* New York: Macmillan, 1923. Pp. 344.
48. SKAGGS, E. B. Personalistic psychology as science. PSYCHOL. REV., 1945, 52, 234–238.
49. SLOANE, E. H. Reductionism. PSYCHOL. REV., 1945, 52, 214–223.
50. SPAULDING, E. G. *Beiträge zur Kritik des psychophysischen Parallelismus vom Standpunkte der Energetik.* Halle: Niemeyer, 1900.
51. STRONG, E. W. *Procedures and metaphysics.* Berkeley: Univ. of California Press, 1936. Pp. 301.
52. VANSITTART, R. G. *Bones of contention.* New York: Knopf, 1945. Pp. 158.
53. VORONKOV, I. M. *Kurs teoreticheskoĭ mekhaniki.* Moscow: State Publ. House of Technico-Theoretical Literature, 1944. Pp. 436.
54. WHELAND, G. W. *The theory of resonance and its application to organic chemistry.* New York: Wiley, 1944. Pp. 316.
55. WINDELBAND, W. Geschichte der Philosophie. In *Die Philosophie im Beginn des zwanzigsten Jahrhunderts* (W. Windelband, Ed.). Heidelberg: Winter, 1907, pp. 529–554.
56. ——. Geschichte und Naturwissenschaft. In *Das Stiftungsfest der Kaiser-Wilhelms-Universität.* Strassburg: Heitz und Mündel, 1894, pp. 15–41.
57. ——. *Introduction to philosophy.* London: Unwin, 1921. Pp. 365.
58. WOLF, A. Recent and contemporary philosophy. In *An outline of modern knowledge.* New York: Putnam, 1931, pp. 542–594.
59. ZHEBRAK, A. R. Soviet biology. *Science,* 1945, 102, 357–358.

A NOTE ON THE SIMILARITIES AND DIFFERENCES BETWEEN PROJECTIVE TESTS AND ABILITY TESTS

BY DANIEL J. LEVINSON

University of California

The recent trend toward widespread use of projective techniques in personality research and clinical therapy has led to considerable controversy concerning the scientific objectivity and validity of such methods. There are many psychologists who accept ability tests (aptitudes, intelligence, etc.) and hold them up as a model and who in comparison reject the projective tests (Thematic Apperception, Play, Rorschach, etc.). Even supporters of the latter tests are often inclined to regard them as categorically different from the former. The writer feels, however, that the similarities between the two types of test are at least as important as their differences; whether a psychologist emphasizes the former or the latter depends on his company at the moment.

This paper is an attempt to distinguish and clarify the underlying similarities and differences between the two sets of tests. It will be held that they are *similar methodologically*, that is, with respect to the situational controls set up, the general process of inference and measurement and, in the most general sense, with respect to the kinds of determiners inferred and measured. They are thought to *differ interpretively*, that is, with respect to the test-relevant dimensions of behavior and the specific nature of the determiners inferred and measured. These points will be discussed within the framework of the general test methodology common to both tests.

I. The Standardized Situation

In both types of test the situation is standardized so that it will be identical for all subjects. This is necessary in order that individual differences in behavior can be attributed primarily to intra-personal characteristics of the subjects (what they bring with them to the experimental-test situation, e.g., IQ, mechanical aptitude, dependency drive). That is to say the difference between the scores of two subjects must be related to a difference in their internal characteristics. In order to achieve this the situation has to be standardized at least with regard to the following aspects:

(1) The *materials* (numbers to be memorized, mazes to be run through, pictures to which stories must be told) are identical for all subjects. Furthermore, no matter what the test may be, the materials must have at least the following characteristics: (a) They must be *interesting* for all subjects. This factor tends to complicate all types of tests; not only do individual differences exist in any small sample in the degree of interest aroused by the test materials, but there are undoubtedly variations over a period of years and in different cultural regions in the 'average interest value' of a given type of material. There are many amusing anecdotes about the inapplicability of the Stanford-Binet materials to Kentucky mountainfolk or Indians because of low interest or distorted meanings. Similarly, certain pictures in the The-

matic Apperception Test, or certain toys in the projective play technique, ought probably to be changed or interpreted differently in different cultures or in the same culture in different historical periods. For instance, the interpretation of play with a 'soldier' or 'tank' toy in a war period would be somewhat different from the interpretation of similar play in a peace period. This is certainly not an insurmountable problem; the point is that it is a problem common to all tests. (b) The materials must present a *problem to be worked out* in terms of the test-relevant intra-personal characteristics. In the ability tests the material offers a difficult problem of a logical, motor, mathematical or other aptitude nature. In the projective tests the problem, though seemingly only perceptual or imaginal or intellectual, is worked out largely in terms of personality-motivational patterns. It has often been said that the projective materials are 'meaningless'; this, in the writer's opinion, is an error. Rather, it should be said that the intrinsically contained meanings (and there are such!) are *inadequate* for solution of the problem; the subject, in working out the problem, tends to select and especially to impose additional meanings in terms of his personality trends. This is similar to the ability test, in which the subject imposes some order and meaning on the initially inadequate material in terms of those inner characteristics (namely, abilities) relevant to the problem. (c) The materials are so selected as to produce a *maximum of individual differences in the relevant behavior dimensions* (number of items answered correctly, kinds of stories told, etc.). The individual differences are harder to quantify in the projective tests because of their greater qualitative complexity (see below, section B).[1] Nevertheless, both tests have in common the requirement of obtaining maximal individual differences. In the ability tests this requirement takes the simpler form of a maximum of quantitative differences on one or a few dimensions (various aptitudes or various 'components' of 'intelligence'). It is usually accomplished by the use of a series of problems of such difficulty that a large range of scores (distribution of successes and failures) will occur. The 'difficulty' of the projective test material results in greater qualitative variations in behavior, and most of these variations are used by the tester for inference and measurement.

(2) The tests must also be standardized with respect to the *physical situation*. It must be comfortable and pleasant, with a minimum of distractions; in short, it must be generally optimal for test behavior.

(3) All tests require uniformity with respect to the *relation of S to E*. E should show much friendliness and encouragement and attempt to establish maximal rapport. But E is *neutral* in all ways that might directly affect the dimensions of behavior relevant to the determiners inferred and measured by the test. That is to say individual differences in 'scores' (quantitative num-

[1] From the point of view of the therapist the complexity and richness of projective test behavior is a rather unmixed advantage. Some academic researchers, however, would prefer to obtain quantitative differences on only a few dimensions which might then be used for rigorous experimental-statistical research. This limitation of behavior dimensions could be imposed by limiting the possible behavioral variability (e.g., by presenting more highly structured material) or simply by ignoring all but a few dimensions. The question of the desirability of such limitations is similar to other historical psychological questions: reliability *vs.* validity, precise narrow measurement *vs.* intuitive broad denotation, molar *vs.* molecular approach, etc.

bers or qualitative categories not yet quantified) must not be attributable to E's influence. In the case of the Stanford-Binet, for example, many a novice tester must be especially careful not to suggest the answer to a test item, not to participate directly in S's problem-solving attempts. Similarly, the projective tester must learn to avoid asking leading questions and to take only the most passive role (though at the same time permissive and supporting) while S is solving the problem which has been set. In both situations E attempts to create an emotionally optimal working atmosphere and at the same time to be neutral and non-participative with regard to S's immediate task.

(4) *S's relation to the task* should be standardized at an optimal level. A maximum of interest and effort should be evoked by the materials, the general situation and E. All tests attempt to obtain some ego-involvement of S, some feeling that the task is important and to-be-accepted. Although the ability tests are usually regarded by S as 'work,' it is important that he have a positive relation to the task and a desire to give an effortful performance.

(5) *The control of S's behavior* must also be regulated in all test situations. S must be allowed the fullest and freest self-expression of *those tendencies* (abilities, needs or whatever) *to be measured by the test*. There must be a minimum of control of the relevant dimensions of behavior.

II. The Process of Inference and Measurement of Intra-Personal Characteristics

All tests have in common the fact of *selectivity of cues* (dimensions of the subject's behavior) utilized for inference and measurement by the psychologist.[2] In ability tests the main behavior dimension used is 'correctness,' and values on this dimension, are measured in terms of time taken, number of errors or successes and/or partial successes, and so on. Responses are categorized in terms of the 'logic of the problem.' Items may be constructed so that only a few alternatives of response exist (true-false, multiple choice), or the possible variations in response may be more subtle and qualitatively complex, thus requiring some type of rating method (e.g., essay questions, vocabulary items on the Stanford-Binet). In any case, the 'score' derived from the test is based on a selection of one part of the total test behavior, namely the end-product of the motor or intellectual performance. Furthermore, only certain aspects of this segment of the test behavior are considered, namely those related to its correctness in solving the problem which has been set.

In the projective tests as well there is selectivity of behavior dimensions to be scored, as has so often been emphasized (as if it were a new phenomenon in the test method). Sometimes the projective tester is concerned with behavior features referring to intellectual level ('form level' in Rorschach, 'constructiveness' in Play), a variable similar to that obtained from an ability test. But in general other types of interpretive cues are used: affect, individuality in selection of material (acceptance, emphasis, rejection and avoidance) and in ways of handling the material, the content and sequence of behavior; the relation to E and to the task, etc.

Another common problem in test method is that of *scoring*. Both ability

[2] This use of the term 'cues' may be somewhat objectionable to the experimental psychologist, to whom it means 'perceptual cues' used by the human or animal subject. It nevertheless seems appropriate in this new context, especially since it emphasizes the selective nature of the psychologist's perception and recording.

tests and projective tests now permit some quantification. The greater precision of measurement (smaller and more reliable scoring units) of the ability tests has been gained only after much research. The projective tests will probably require more time, effort and thought, but there is no intrinsic reason that they cannot be developed to a point of equal precision.

In comparing the two types of tests, a number of added difficulties in the statistical development of the projective tests must be mentioned. (a) More complex behavior is elicited in the projective test; the task as presented permits the widest expression of S's creative tendencies. (b) There is a greater number of relevant behavior dimensions and the complexity of the inferences (patterns of scores) is greater. (c) More behavior dimensions must be used in determining a specific score on one variable; this raises the problem of the rating method which, as stated above, is common to a great many tests. The more complex the rating task, the more training must scorers be given; but the fact that a rating method is used is not an intrinsic weakness of the test, and there are adequate statistical techniques (reliability) for evaluating obtained ratings. (d) The statistical treatment of the scores obtained is complicated by the fact that the form of their distribution often deviates widely from the normal curve. Also, statistical techniques for handling many variables at a time are as yet rather inadequate. (e) The 'summation hypothesis' (that a person's standing on a given dimension equals the sum of his item or sub-unit scores) does not always hold for personality variables. (It has not, for that matter, been proved to hold for intelligence either, but much useful research has been based on that assumption.) In some cases the fact that a person is high on one item (or shows a given characteristic to a marked degree) and low on another item is used to give him a higher total score than he would get by being high on both items. The introduction of clinical reasoning into testing methodology is still at an intuitive, preprecision stage, but it is certainly not over-optimistic to believe that more formalized and rigorous rules for getting scores on significant variables can be worked out and that adequate statistical treatment can be obtained.

III. General Properties of Test-Inferred Variables

What the ability tests measure is obviously different from that which is measured by projective tests. However, it seems worth emphasizing that the inferences to which the two sets of tests lead (that a person has a certain amount of a certain attribute), though different in content, nevertheless have certain general properties in common: they all refer to *intra-personal characteristics*, that is, to *relatively stable psychological 'dispositions,'* tendencies or readinesses which the individual brings with him to the situation and which are the primary determiners of individual differences in the test-relevant behavior dimensions.

The idea that individual differences on the 'ability' dimensions are *determined* has been widely accepted. When one subject exhibits more of a certain kind of behavior than another (e.g., gets more items correct) he is said to be more 'intelligent' or to have more 'mechanical aptitude.' Several general properties of this inference must be noted: (a) An internal ability or 'tendency to succeed' is regarded as a determiner of the behavior difference, that is, the behavior and the ability are re-

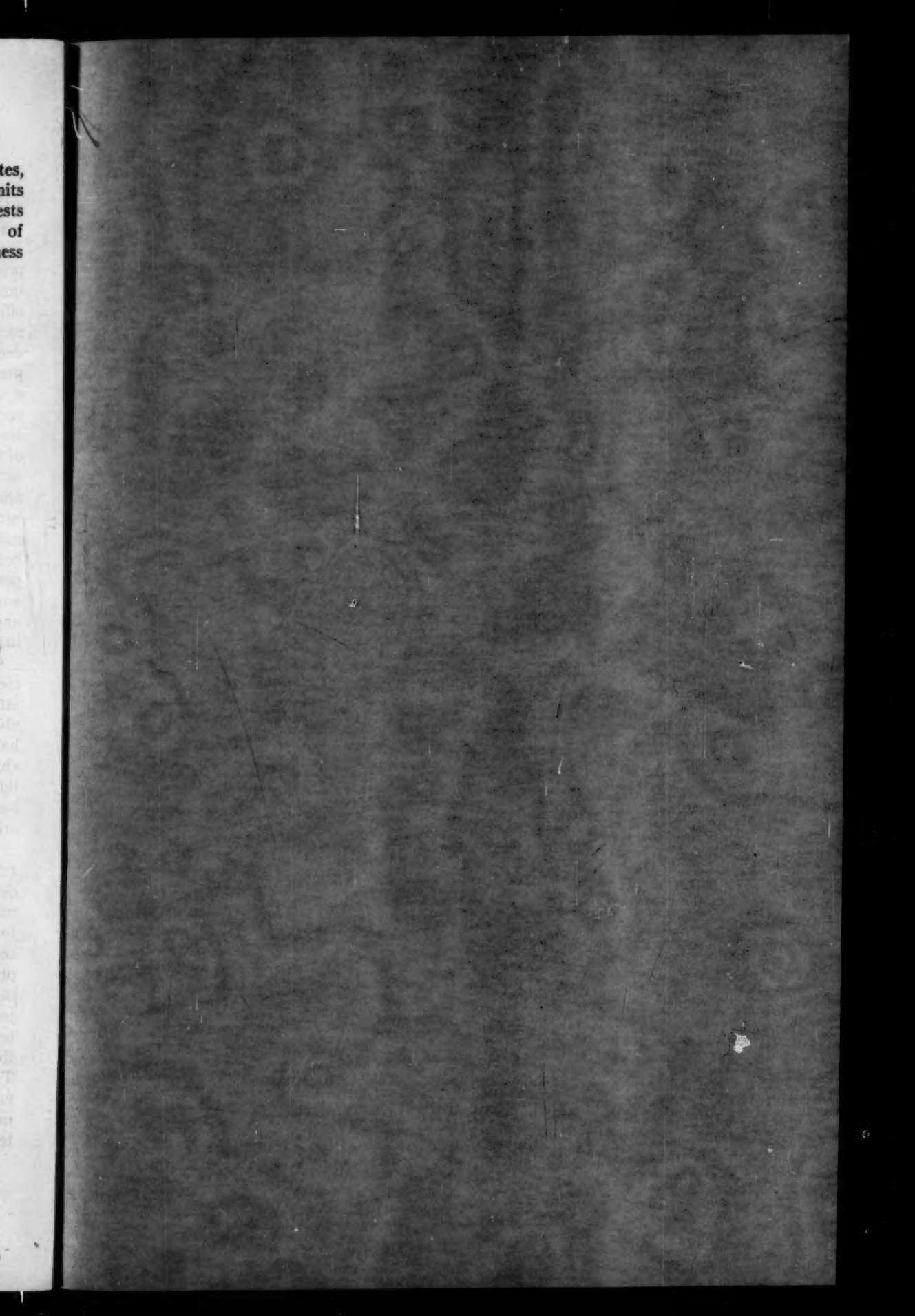
garded as *lawfully connected.* (b) The disposition is thought to have a certain *general validity;* it supposedly extends beyond the test situation; IQ, for example, may be used as a criterion for predicting behavior in situations as distant from the test as 'being an army officer' or 'being a government employee.' In other words, the same underlying tendency is conceived as expressing itself in a variety of forms in a variety of situations. (c) The tendency is regarded as having a certain *stability.* Regardless of the question of the genetic determiners of the ability as tested at a given time—we may ignore the nature-nurture and other genetic problems—most investigators regard the abilities usually tested as being relatively difficult to change (especially in the adult), limited in the amount of change that can be induced, and subject to 'laws of change' (learning, unlearning, transfer, etc.).

To summarize, the abilities (tendencies to succeed) are usually regarded as internalized, relatively stable, measurable, underlying determiners of behavior which are subject to laws of change, have wide significance (validity) and are expressed in different but recognizable forms in different situations.

The 'dispositions' of personality (drives, conflicts, defenses, etc.) are derived from the projective test in a manner comparable to that described for the ability tests, and they are conceived as having the same general properties. (a) Individual differences in the 'personality' dimensions of projective test behavior are related to (determined by) individual differences in the strength of personality tendencies. That is to say behavior differences are not due primarily to 'chance' or 'momentary interest' factors; they are determined by internal dispositions as fully and as lawfully as are differences in the 'ability' dimensions of behavior in the ability test situation. (b) The tendencies (which are now or someday will be) measured are thought to have underlying importance in the personality and thus to find wide expression in many areas of the person's life. It is thought that in a situation offering a maximum of encouragement, permissiveness and variety of behavioral possibilities, those dynamic trends will emerge which have great personality significance. The scores obtained on a projective test should be relevant, no less than the IQ score, to such far-removed situations as the home or the job. (c) The personality tendencies are regarded as having a certain stability in the adult, as being difficult to change, and as subject to 'laws of change' (insight, learning, sublimation, channelization, etc.) which, though somewhat different in nature, are not different in principle from the laws of change of abilities.

IV. SUMMARY

Emphasis in the discussion above has been on the similarities of projective tests and ability tests with respect to certain general characteristics of their experimental, statistical and conceptual methodology. Emphasis on the similarities should not obscure the many important differences between the two types of method. Some of these differences, as brought out above, are: the nature of the relevant behavior dimensions, the kinds of tendencies inferred and measured, the complexity of the cues (behavior dimensions) used in measuring a given variable, the complexity of the variable patterns, and the greater indirectness with which personality tendencies are expressed. Much further work is required not only in the use of projective tests but also in mak-

ing explicit the specific methodological steps involved.

The attempt has been made to demonstrate that projective tests are similar to ability tests in experimental-conceptual methodology. The scientific validity of this methodology indicates, it is believed, that the ultimate limits of the development of projective tests are set primarily by the ingenuity of their users and not by any weakness inherent in the nature of such tests.